



Facultad de Ingeniería y Computación
Escuela Profesional de Ingeniería de Telecomunicaciones

**“Diseño de un estimador no lineal para predecir
el nivel de nitrógeno en suelo agrícola”**

Presentado por:

Katty Miluska Sánchez Mora

Para Optar por el Título Profesional de:

INGENIERÍA DE TELECOMUNICACIONES

Orientador: “Efraín Tito Mayhua López”

Arequipa, septiembre de 2018

**PROGRAMA PROFESIONAL DE INGENIERÍA DE
TELECOMUNICACIONES**

**Diseño de un estimador no lineal para predecir el nivel de
nitrógeno en suelo agrícola**

Autor: Katty Miluska Sánchez Mora

Septiembre, 2018

“Para conseguir cosas importantes, se necesitan dos cosas: un plan, y no
abandonar antes de tiempo”

Leonard Bernstein

Índice general

Abstract	3
Resumen	4
1 Introducción	5
1.1 Motivación y Contexto	6
1.2 Planteamiento del problema	6
1.3 Objetivos	7
1.3.1 Objetivo general	7
1.3.2 Objetivos específicos	7
1.4 Organización del Presente Documento	7
2 Marco Teórico	9
2.1 Importancia del nitrógeno en el suelo agrícola	9
2.1.1 Pérdidas de nitrógeno	9
2.1.2 Uso eficiente del nitrógeno en sistemas de cultivo	11
2.1.3 Cálculo del nivel de nitrógeno del suelo para la agricultura de precisión por métodos directos	11
2.1.4 Cálculo del nivel de nitrógeno del suelo para la agricultura de precisión por métodos indirectos	15
2.2 Aprendizaje máquina para estimación	15
2.2.1 El problema de regresión	15
2.2.2 Generalización y sobre-ajuste	17
2.2.3 Entrenamiento con pocas muestras	18
3 Estado del Arte	20
3.1 Muestreo del suelo basado en microdiálisis	20
3.2 Método de electrodos de ión selectivos	22
3.2.1 Electrodos selectivos de iones de nitrato — in-situ	22
3.2.2 Electrodos selectivos de iones in-situ en tiempo real	24
3.3 Red de sensores in-situ	26
3.3.1 Muestreo anidado para la medida de variabilidad espacial	27
3.3.2 Optimización de una red de sensores dentro de campo para el control de nitrógeno soluble en el suelo	28
3.4 Otros sensores	29
3.4.1 Nano sensores electroquímicos	29
3.4.2 Nano sensores de electrodos de alambre	29

3.4.3	Biosensores	29
3.4.4	Tecnología de laboratorio sobre un chip	29
4	Propuesta de Solución	31
4.1	Toma de muestras de las condiciones ambientales del suelo	31
4.2	Toma de muestras de los niveles de nitrógeno en suelo	32
4.3	Análisis de las muestras	33
4.4	Diseño de un modelo máquina para la estimación de los niveles de nitrógeno en suelo agrícola	35
5	Experimentos y Resultados	39
5.1	Configuración de los experimentos	39
5.2	Experimentos	39
5.3	Resultados	42
5.4	Discusión	43
6	Conclusiones y trabajos futuros	47
6.1	Conclusiones	47
6.2	Publicaciones generadas	48
6.3	Trabajos futuros	48
	Acknowledgements	49
	Bibliografía	50
	Nomenclatura	52

Índice de figuras

2.1.	Rango de nitrógeno recomendado para cada cultivo.	13
2.2.	Diagrama de los procesos del método de Kjeldahl.	14
3.1.	Diagrama de una sonda de microdiálisis CMA 20 [Shaw, 2015a]. . .	21
3.2.	Representación esquemática de un sistema ISE, mostrando que el ISE y el electrodo de referencia son enlazados por un voltímetro [Shaw, 2015a].	23
3.3.	Diseño de muestreo optimizado de una estación principal [Shaw, 2015b].	27
3.4.	Concentración de nitrógeno del suelo para cada incremento de profundidad de 5 cm [Shaw, 2015b].	28
4.1.	Despliegue de sensores en la plataforma inteligente para el sistema riego.	32
4.2.	Muestras de tierra agrícola etiquetadas para su análisis de nitrógeno en el laboratorio.	33
4.3.	Condiciones ambientales de temperatura, conductividad eléctrica, humedad y nitratos medidos en los terrenos cultivado con: (a) paprika y (b) alcachofa.	36
5.1.	Diagrama de dispersión de los centros seleccionados para el modelo estimador de: (a) paprika y (b) alcachofa.	40
5.2.	Valores del entrenamiento del modelo máquina para la paprika con diferentes variables de entrada. (a) : con las variables de entrada sin normalizar y (b) : con las variables de entrada normalizadas.	41
5.3.	Valores del entrenamiento del modelo máquina para la alcachofa con diferentes variables de entrada. (a) : con las variables de entrada sin normalizar y (b) : con las variables de entrada normalizadas.	42
5.4.	Resultados del estimador del nivel de nitrógeno con los datos de prueba para diferentes parámetros de entrada para el cultivo de paprika. (a) : con parámetros sin normalizar y (b) : con parámetros normalizados.	44
5.5.	Resultados del estimador del nivel de nitrógeno con los datos de prueba para diferentes parámetros de entrada para el cultivo de alcachofa. (a) : con parámetros sin normalizar y (b) : con parámetros normalizados.	45

Índice de tablas

2.1. Niveles de nitrógeno recomendados para los cultivos [Kissel and Sonon, 2008].	12
5.1. Promedio del valor absoluto del error y desviación estándar del esti- mador en el entrenamiento.	43
5.2. Promedio del valor absoluto del error y desviación estándar de los niveles estimados para los datos de prueba.	46

Abstract

The loss of macronutrients and micronutrients from agricultural land is one of the processes that are always present in agriculture. Thus, is necessary use of fertilizers NPK (Nitrogen, Phosphorus and Potassium) to optimize crop yields, increase profitability and minimize losses.

Among the NPK fertilizers, nitrogen is the essential nutrient in the process of plant growth. Thus, it is that in agriculture nitrogen is applied to plants during irrigation, so care should be taken in the amounts of water that has been granted to the crop, as its excess leads to a loss of nitrogen and other nutrients by leaching.

Given this reality, technology capable of monitoring and measure the Nitrogen levels in soil locally and in real time, so that in this way the necessary amount of nutrients can be given to the crop.

In recent years there has been proposed a number of techniques and methods (direct and indirect) for measurement of nitrogen. Direct methods can be in-situ or in laboratory, but they tend to be really expensive and/or dependent on soil conditions. Instead indirect methods, can estimate Nitrogen levels in-situ and in real time, based on measuring other parameters, at the expense of accuracy.

Weighing the advantages and disadvantages, is that an indirect method was designed, based on algorithms own from the machine learning area (ML), it will be capable of predicting future values of the Nitrogen levels in soil, after learning a model from a set of observations. These observations are given by parameters such as electrical conductivity, temperature and soil moisture. The values of these parameters were acquired through sensors that exist in the market. For the validation of the method, experimental tests were carried out with real data that were measured in the Santa Gabriela S.A.C. of Santa Rita de Siguan, Arequipa, Peru.

From the results obtained from the estimator it has been noted that the most influential variables to obtain the estimated nitrogen levels are the electrical conductivity and temperature, which gives reasonable results within a confidence interval of 99% given by the range of permitted levels for the crop.

Keywords: nitrogen, agricultural soil, estimator, machine learning.

Resumen

La pérdida de los macronutrientes y micronutrientes del suelo agrícola es uno de los procesos que siempre están presentes en la agricultura. Por estas pérdidas se hace necesario el uso de fertilizantes NPK (Nitrógeno, Fósforo y Potasio), para optimizar los rendimientos del cultivo, aumentar la rentabilidad y reducir al mínimo las pérdidas.

Entre los fertilizantes NPK, el nitrógeno es el nutriente esencial en el proceso de crecimiento de la planta. Por ello, es que en la agricultura el nitrógeno es aplicado a las plantas durante el riego, por lo que se debe tener cuidado en las cantidades de agua otorgada al cultivo, ya que su exceso conlleva a una pérdida del nitrógeno y otros nutrientes por la lixiviación.

Ante esta realidad, se requiere de tecnologías que permitan monitorear y medir los niveles de nitrógeno presentes en el suelo agrícola in-situ y en tiempo real, para que de este modo se pueda otorgar la cantidad necesaria de nutrientes al cultivo.

En los últimos años se han propuesto una serie de técnicas y métodos (directos e indirectos) para la medida del nitrógeno. Los métodos directos se pueden realizar en área de cultivo o en laboratorio, pero resultan ser muy caros y/o dependientes a las condiciones del suelo. En cambio, los métodos indirectos pueden estimar los niveles de nitrógeno in-situ y en tiempo real basados en medir otros parámetros, a costa de precisión.

Sopesando las ventajas y desventajas, es que se diseñó un método indirecto, basado en algoritmos propios del área del Aprendizaje Máquina (ML), que será capaz de estimar valores de los niveles de nitrógeno en el suelo agrícola, tras aprender un modelo a partir de un conjunto de observaciones. Estas observaciones están dadas por parámetros tales como la conductividad eléctrica, temperatura y humedad del suelo. Los valores de dichos parámetros son adquiridos a través de sensores que existen en el mercado. Para la validación del método se realizaron pruebas experimentales con datos reales que fueron medidos en el fundo Santa Gabriela S.A.C. de Santa Rita de Sigüas, Arequipa, Perú.

De los resultados obtenidos del estimador se ha podido notar que las variables más influyentes para obtener los niveles de nitrógeno estimados son la conductividad eléctrica y la temperatura, lo que otorga resultados razonables dentro de un intervalo de confianza del 99 % dados por el rango de los niveles permitidos para el cultivo.

Palabras clave: nitrógeno, suelo agrícola, estimador, aprendizaje máquina.

1 Introducción

En la agricultura los macronutrientes y micronutrientes del suelo pueden perderse, y no estar disponibles para las plantas. La disminución en estos niveles de nutrientes naturales del suelo, tradicionalmente, se compensan con el uso de fertilizantes tales como el Nitrógeno (N), Potasio (K) y Fósforo (P), cuya combinación se denomina NPK.

Los fertilizantes NPK son necesarios para el crecimiento de las plantaciones en los distintos cultivos [Gaikwad and Galande, 2015]. El uso combinado de estos fertilizantes juega un papel importante en la producción agrícola; su aplicación de forma equilibrada y en el momento adecuado tiene un gran impacto en el rendimiento de los cultivos [Malghani et al., 2010]. El Fósforo estimula el florecimiento y la formación de la semilla, el Potasio es también importante pues es el encargado de la activación de varias enzimas, de la mejora de las proteínas, hidratos de carbono y la concentración de grasa, también desarrolla la tolerancia contra la sequía además de la resistencia a las heladas, y el ataque de enfermedades. El Nitrógeno es el nutriente esencial por ser vital en el proceso de crecimiento de las plantas, ya que es contribuyente de la clorofila, proteínas y ácidos nucleicos; además, es un elemento central por su papel en la síntesis de sustancias [Malghani et al., 2010].

La absorción de los nutrientes en las plantas se realiza a través de las raíces y las hojas. Las hojas absorben carbono, oxígeno del aire y también unas pequeñas cantidades de nutrientes, y las raíces absorben el agua del suelo en el que van disueltos la mayor cantidad de los nutrientes. El agua del riego se infiltra en el suelo y va llenando los poros de las partículas del suelo. Al llenarse los poros de agua, parte de esta cae hacia abajo provocando la lixiviación de los nutrientes. Es por ello que el riego excesivo causa que los nutrientes suministrados no sean aprovechados por las plantas y al contrario se den por perdidos.

Debido a este inconveniente, es necesario conocer los niveles de nitrógeno presentes en el suelo durante el riego para determinar si los nutrientes están sufriendo pérdidas causadas por la lixiviación, desnitrificación y volatilización.

Existen varios métodos para medir el nivel de nitrógeno en el suelo. Un método es tomar muestras del suelo para analizarlo en el laboratorio, lo que resulta ser caro y además requiere de la presencia y conocimientos de un personal. Otro de los métodos es utilizar un modelo no lineal con Kernels Gaussianos basados en técnicas de aprendizaje máquina que hace posible estimar un comportamiento a partir de variables empíricas que pueden ser las condiciones ambientales del suelo como la conductividad eléctrica, temperatura y humedad.

1.1. Motivación y Contexto

La pérdida de nutrientes en el suelo agrícola es un proceso presente en la agricultura, por lo que los agricultores deben suministrar dichos nutrientes a través del riego. El riego excesivo provoca que los nutrientes sean lixiviados y como consecuencia no son absorbidos por las plantas. Por ello es necesario conocer los niveles de nitrógeno presentes en el suelo, para saber si la cantidad suministrada es la necesaria para que las plantas puedan crecer y desarrollarse de forma adecuada.

En los últimos años se han propuesto una serie de técnicas y métodos para medir el nitrógeno, los cuales pueden ser clasificados en métodos directos e indirectos. Los métodos directos que se utilizan actualmente pueden ser in-situ o en laboratorio, pero en ambos casos, requiere de solutos químicos. Los métodos indirectos se basan en estimar los niveles de nitrógeno en el suelo a partir de medir otros parámetros del suelo. Entre los métodos directos e indirectos más utilizados podemos mencionar a la microdiálisis y el de uso de electrodos de ión selectivo. Estos métodos toman muestras en el área de cultivo y posteriormente son llevadas al laboratorio para ser analizadas. Por lo tanto, el análisis no es en tiempo real, resultando en un método poco eficiente. Además, el transporte y almacenamiento de las muestras antes de ser analizadas en el laboratorio conduce a un mal muestreo, ya que se pueden producir cambios químicos en las muestras del suelo [Shaw, 2015a]. Por todos estos motivos, se hace necesario el desarrollo de nuevos procedimientos para medir el nivel de nitrógeno en tiempo real y de manera automática. Por otro lado, el aprendizaje máquina, son un conjunto de técnicas computacionales que a partir de un conjunto de observaciones del problema (medidas) aprende un modelo para estimar valores de la variable observada.

En este trabajo de tesis se propone el uso de algoritmos de aprendizaje máquina (ML, del inglés *Machine Learning*) para diseñar un modelo que estime los niveles de nitrógeno presente en el suelo, pues hasta el momento las ventajas del ML no han sido aprovechadas y no se han aplicado ni desarrollado para solucionar este tipo de problemas.

1.2. Planteamiento del problema

El principal inconveniente en la producción agrícola es conocer si la cantidad de nitrógeno presente en el suelo es la adecuada para el desarrollo y crecimiento de las plantas.

Las técnicas existentes para la medida del nivel de nitrógeno en el área de cultivo no son prácticas, pues se necesita que el personal tenga conocimiento para la toma de medidas y que esté presente en campo para instalar la instrumentación de medida que en la mayoría de los casos está formada por electrodos los que durante el proceso de análisis son ensuciados lo que conlleva a un error en la toma de medidas, además

la disponibilidad del personal conlleva a un aumento de costos y una inversión de tiempos adicionales.

Otro inconveniente con los métodos in-situ existentes, es que el transporte y almacenamiento de las muestras para ser analizadas en el laboratorio, conduce a un mal muestreo porque pueden producir cambios químicos en las muestras del suelo.

1.3. Objetivos

1.3.1. Objetivo general

El objetivo principal de este trabajo de tesis es diseñar un estimador no lineal para estimar el nivel de nitrógeno en el suelo agrícola. Para ello, vamos a utilizar técnicas del aprendizaje máquina que nos permitirán desarrollar un método de medida indirecta y en tiempo real de los niveles de nitrógeno en suelo a partir de observar los niveles de la conductividad eléctrica, temperatura y humedad del suelo.

1.3.2. Objetivos específicos

Para lograr el objetivo general de este proyecto, se tiene que llevar a cabo los siguientes objetivos específicos:

1. Realizar mediciones de las condiciones ambientales y niveles de nitrógeno en suelo agrícola.
2. Seleccionar las características ambientales del suelo que afectan el comportamiento de los nutrientes suministrados al suelo agrícola.
3. Diseñar un modelo máquina para la estimación de los niveles de nitrógeno en suelo agrícola.
4. Entrenar y validar el modelo máquina diseñado.

1.4. Organización del Presente Documento

En el segundo capítulo, se presenta el marco teórico que explica la teoría y los conceptos necesarios para poder comprender los términos del proyecto. Los temas tratados son:

- Importancia del nitrógeno en el suelo agrícola, donde se explican los procesos de pérdidas de nitrógeno en el suelo.
- La importancia de hacer un uso eficiente del nitrógeno en los sistemas de cultivo, ya que el uso excesivo o deficiente del fertilizante puede afectar la calidad del cultivo, por lo que se recomienda las dosis de nitrógeno que deben ser aplicadas.

- Cálculo de los niveles de nitrógeno en el suelo agrícola para obtener una agricultura de precisión, esto se puede lograr a través de métodos directos, como pruebas rápidas en el campo y métodos de Kjeldahl. Estos son llamados métodos directos porque realizan el análisis del suelo a través de solutos químicos. También se puede lograr por métodos indirectos que se basan en estimar los niveles de nitrógeno a partir de medir otros parámetros del suelo, esto se puede lograr por la instalación de sensores en el área de cultivo y adquisición de datos en tiempo real.
- Cómo estimar un modelo por aprendizaje máquina a través de la regresión lineal. También se explica cómo se puede obtener un estimador. Los parámetros de dicho modelo son hallados minimizando una función de coste.
- Se explica los conceptos de generalización, sobre ajuste y validación cruzada.
- Como trabajar el entrenamiento con pocas muestras.

En el tercer capítulo, se presenta el estado del arte, donde se da a conocer los antecedentes e investigaciones recientes que nos van a ayudar a tener un enfoque acerca de los métodos utilizados para medir los niveles de nitrógeno. Aquí se habla de las ventajas y desventajas de los diferentes métodos realizados únicamente in-situ, y otros realizados in-situ y en tiempo real.

En el cuarto capítulo, se explica la distribución para la toma de muestras de las condiciones ambientales del suelo, como se tomaron y analizaron las muestras para conocer los niveles de nitrógeno en el suelo y se explica la propuesta para realizar el diseño del algoritmo utilizado para alcanzar el objetivo propuesto en el trabajo de tesis. Para ello se eligió un modelo máquina no lineal, basado en una expansión gaussiana.

En el quinto capítulo, se presenta los experimentos realizados para el entrenamiento con las diferentes variables de conductividad eléctrica, temperatura y humedad. También encontrará los resultados del modelo estimador de nitrógeno presente en el suelo agrícola para los datos de prueba y su discusión.

En el sexto capítulo, se presentan las principales conclusiones del trabajo de tesis, así como las recomendaciones para futuros trabajos que se puedan realizar en la misma área.

2 Marco Teórico

2.1. Importancia del nitrógeno en el suelo agrícola

El nitrógeno es uno de los macronutrientes esenciales y más importantes para el desarrollo de las plantas. La gestión correcta del nitrógeno puede optimizar los rendimientos del cultivo, aumentar la rentabilidad y reducir al mínimo las pérdidas [Shaw, 2015a, Cavigelli et al., 2004].

Cuando se utilizan sistemas de riego tecnificados, como el riego por goteo o fertirriego, la forma de otorgar nitrógeno al suelo es por la misma línea de riego. Por tanto, la gestión del riego es un factor esencial para el crecimiento del cultivo dado que una mala gestión, no solo provoca la falta o exceso de agua, sino, la pérdida o exceso de nutrientes, entre ellos el nitrógeno como elemento esencial para el desarrollo del cultivo.

2.1.1. Pérdidas de nitrógeno

Son tres los procesos de pérdida de nitrógeno: desnitrificación, volatilización y lixiviación, siendo los más importantes la lixiviación y la desnitrificación, que ocurren por exceso de agua en el suelo.

Pérdida por lixiviación

El proceso de lixiviación o lavado es causado por las lluvias y el riego, dado que el nitrato por su carga negativa se mueve fácilmente hacia abajo junto con el flujo de agua, debido a que el nitrato no es retenido por las partículas del suelo porque ambos están cargados negativamente [Perdomo and Barbazán, 2011, Johnson et al., 2005].

Los factores que afectan las pérdidas por lixiviación son:

- **Tipo de suelo:** El tipo de suelo afecta la magnitud de las pérdidas por lixiviación, siendo mayor cuando el suelo es liviano, es decir de partículas grandes.
- **Humedad del suelo:** Es posible que las pérdidas sean mayores en suelos relativamente secos.
- **Presencia de rastrojo en superficie:** El rastrojo aumenta la cantidad de agua que infiltra en relación con la que escurre.

Pérdida por desnitrificación

La desnitrificación ocurre cuando las bacterias utilizan el nitrato como fuente de oxígeno por estar el suelo saturado o cuando el suelo está muy húmedo.

En el proceso las bacterias convierten el nitrato en gas de nitrógeno y al estar en este estado vuelve a la atmósfera [Perdomo and Barbazán, 2011].

Los factores que afectan las pérdidas por desnitrificación son:

- **Energía:** El proceso requiere de energía externa, por lo que debe existir en el suelo materia orgánica de fácil descomposición, haciendo que la tasa de desnitrificación tienda a ser mayor en la superficie.
- **Temperatura:** A mayor temperatura la tasa de desnitrificación también aumenta. Se debe considerar que los microorganismos son capaces de adaptarse a cualquier temperatura.

Pérdida por volatilización

La volatilización es la pérdida de nitrógeno que ocurre cuando el amonio se convierte en amoníaco y se libera a la atmósfera.

Las pérdidas aumentan cuando el pH en el suelo es mayor y cuando las condiciones favorecen la evaporación, por ejemplo tener un clima caluroso y con vientos [Johnson et al., 2005, Perdomo and Barbazán, 2011].

Los factores que afectan las pérdidas por volatilización son:

- **Reacción del suelo:** Aumentar los niveles de pH del suelo tiende a aumentar las pérdidas de nitrógeno.
- **Temperatura:** A mayor temperatura mayor pérdida de amoníaco.
- **Concentración de amonio:** Al aumentar el amonio los procesos de volatilización también aumentan.
- **Modo de aplicación del fertilizante:** Si el fertilizante se aplica a la siembra, por lo que queda cubierto por el suelo, no hay pérdidas por volatilización. Si se aplica directamente sobre el cultivo, es decir, que no es cubierta por el suelo, la probabilidad de pérdidas aumenta.
- **Humedad:** No hay pérdidas de amoníaco cuando el suelo está con baja humedad; las pérdidas aumentan cuando el suelo está altamente húmedo y se está secando; y las pérdidas se reducen cuando el suelo está altamente húmedo y no se está secando.

2.1.2. Uso eficiente del nitrógeno en sistemas de cultivo

Para lograr altos rendimientos de producción es necesario fertilizar con nitrógeno, pero también es necesario hacer una fertilización racional, es decir, que la eficiencia del uso de fertilizantes nitrogenados debe ser mejorada para mantener la calidad de la tierra.

2.1.2.1. Deficiencia y exceso de nitrógeno en plantas

Los síntomas de la deficiencia de nitrógeno se expresan en las hojas más viejas. La deficiencia de nitrógeno hace que las cantidades de clorofila se reduzcan haciendo que las hojas se tornen amarillentas (clorosis), las hojas más viejas son las primeras en amarillarse, posteriormente las hojas más jóvenes también lo hacen si la deficiencia de nitrógeno se vuelve severa. Además se presenta un crecimiento detenido o lento, plantas pequeñas, plantas con pocas hojas y rendimiento significativamente reducido [Hassan, 2011].

El exceso de nitrógeno puede afectar la calidad de la planta y el producto, provocando plantas débiles con tejido blando, crecimiento suprimido de la planta, rendimiento de baja calidad, retraso en la producción, reducción del número total de frutas por planta hasta en un 48 a 57 % en cosechas tempranas, además el cuajado de los frutos disminuye linealmente a medida que se incrementan las dosis de nitrógeno. El exceso también provoca que la producción de capullo de la flor se reduzca o que el desarrollo de las flores sea incompleto (sin estambres o sin pistilos), y que la planta posea mayor susceptibilidad a enfermedades y plagas [Hassan, 2011].

2.1.2.2. Recomendación para la aplicación de nutrientes en cultivos

La cantidad aplicada a los cultivos dependerá del rendimiento que se desea obtener y del tipo de cultivo. Es así que en [Kissel and Sonon, 2008] se recomiendan las dosis de nitrógeno que deben ser aplicadas a los diferentes cultivos para su rendimiento óptimo. Estas recomendaciones se basan en experimentos realizados en diferentes cultivos. En la Tabla 2.1 se muestra una relación de varios cultivos con su dosis de nitrógeno recomendado para su aplicación, y en la Figura 2.1 se muestra la representación de dichos rangos.

2.1.3. Cálculo del nivel de nitrógeno del suelo para la agricultura de precisión por métodos directos

2.1.3.1. Pruebas rápidas en el campo

Para las pruebas rápidas del suelo, la extracción de las muestras del suelo se realiza de forma manual, usando una variedad de soluciones, posteriormente se hacen análisis

Cultivo	Nivel Mín. de nitrógeno (ppm)	Nivel Max. de nitrógeno (ppm)
Maíz	9.70	58
Algodón	29	50.70
Manzana	0	24.20
Frambuesa	40	139
Aceitunas	38.60	48.30
Durazno	9.70	14.50
Pera	0	24.20
Pecana	48.30	96.60
Ciruela	9.70	14.50
Fresas	53.10	58
Espárragos	24.20	38.60
Beterraga	43.50	62.80
Brócoli	84.50	108.70
Repollo	84.50	108.70
Melón	48.30	72.50
Zanahoria	48.30	62.80
Coliflor	43.50	53.10
Cebolla china	43.50	60.40
Pepino	38.60	72.50
Berenjena	60.40	84.50
Arvejas	36.20	48.30
Papa	72.50	96.60
Lechuga	60.40	72.50
Cebolla	48.30	72.50
Pimientos	12.10	48.30
Rabanitos	43.50	58
Vainitas	29	48.30

Tabla 2.1: Niveles de nitrógeno recomendados para los cultivos [Kissel and Sonon, 2008].

de nitrato por tiras colorimétricas en combinación con un reflectómetro portátil o electrodos selectivos de ión (ISE, del inglés *ion selective electrodes*).

Aunque este enfoque muestra una mejora, aún se presentan problemas por los límites de precisión y detección [Shaw, 2015a]. Adicional a esto los análisis deben ir acompañados de una determinación del contenido de humedad en el suelo para convertir la concentración de nitrato del suelo. Estas pruebas resultan ser poco probable ya que requiere de gran esfuerzo para obtener datos de alta resolución para informar a la dirección PA.

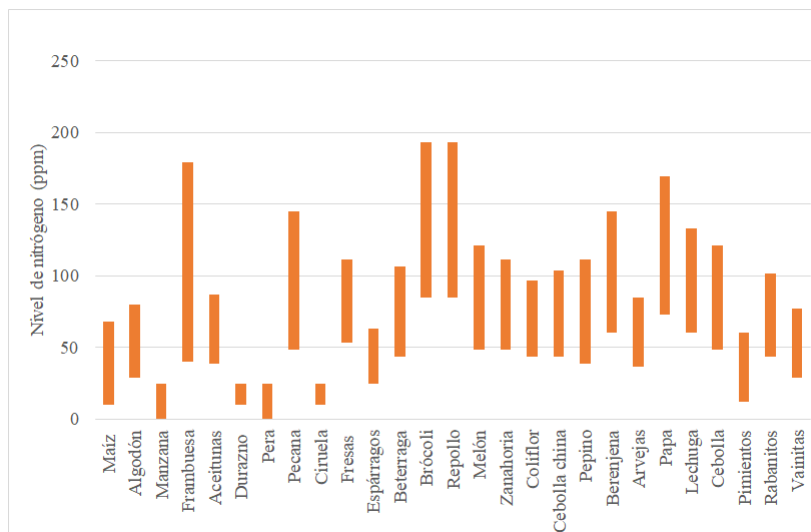


Figura 2.1: Rango de nitrógeno recomendado para cada cultivo.

2.1.3.2. Método Kjeldahl

El método Kjeldahl es un método casi universal y el más antiguo que actualmente sigue siendo usado para determinar el contenido de nitrógeno de sustancias orgánicas e inorgánicas [Thompson et al., 2002, Labconco, 1998]. En realidad, este método utiliza un equipo sencillo y puede ser realizado por un técnico sin experiencia.

El método Kjeldahl se puede dividir en tres etapas principales [Labconco, 1998] que son:

- **Digestión:** La ebullición de una muestra homogénea en ácido sulfúrico concentrado, da como resultado final una solución de sulfato de amonio.
- **Destilación:** Se mezcla la digestión ácida para convertir el amonio a nitrato, seguido por la ebullición y condensación del gas amoníaco en una solución receptora.
- **Valoración:** Cuantificar la cantidad de amoníaco en la solución receptora.

Proceso de medida

Para realizar el análisis de las muestras se utiliza un matraz Kjeldahl que es una bombilla con un cuello condensador posicionado a un lado para el reflujo del ácido. El método utiliza como sustancia base ácido sulfúrico, cuya cantidad depende del tamaño de la muestra, también se debe considerar que la entrada de calor para la mezcla de digestión ácida y la duración de la digestión influye en la pérdida de ácido debido a la vaporización durante el proceso [Labconco, 1998].

El método también adiciona sales para acelerar el proceso de digestión, caso contrario el proceso sería muy largo. La adición de sal inorgánica a la etapa de digestión eleva el punto de ebullición del ácido sulfúrico, cuya temperatura llega a 330°C ; si se utiliza otra sal como el sulfato de potasio (K_2SO_4) la temperatura se eleva a 390°C o más [Labconco, 1998].

Mayores temperaturas elevan la velocidad de descomposición orgánica de las muestras y disminuye el tiempo requerido en el proceso de digestión. Sin embargo, adicionar grandes cantidades de sal puede elevar la temperatura a más de 400°C , lo que puede originar que los compuestos nitrogenados se volatilicen y se puedan perder en la atmósfera.

El matraz Kjeldahl se une a un condensador de agua y se calienta para hervir el gas amoníaco de la digestión como se ve en la Figura 2.2. El extremo del refrigerante se sumerge en un frasco de solución ácida de recepción, ya sea un ácido estándar o ácido bórico, y así con la trampa de Kjeldahl permite la destilación del amoníaco.

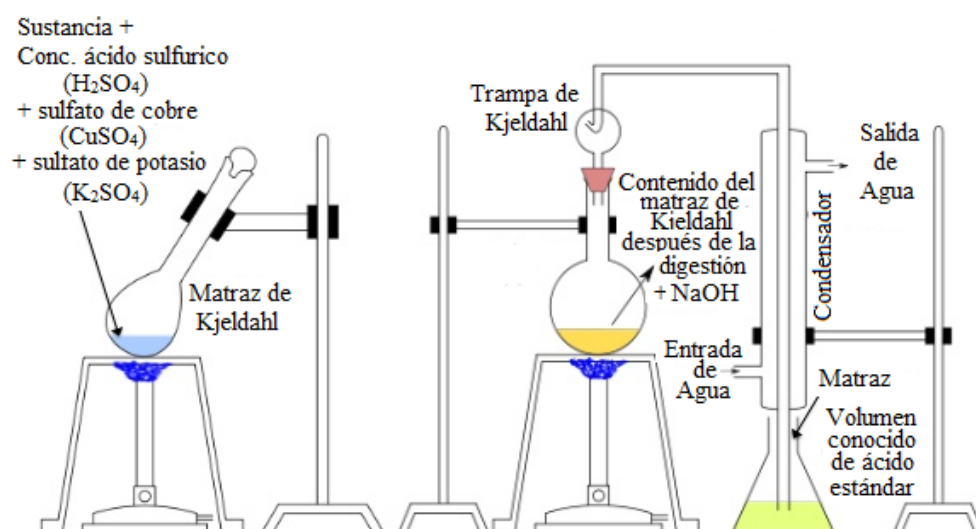


Figura 2.2: Diagrama de los procesos del método de Kjeldahl.

La mayoría de amoníaco es destilado y atrapado en la solución de ácido recibido dentro de los primeros 5 o 10 minutos de ebullición. Los tiempos de destilación y los volúmenes de destilado recogido deben ser estandarizados para todas las muestras.

La tasa de destilación se ve afectado por la capacidad de refrigeración del condensador.

Finalmente, para la determinación del nitrógeno se realiza una valoración la cual indica el amoníaco presente en el destilado con un cambio de color.

2.1.4. Cálculo del nivel de nitrógeno del suelo para la agricultura de precisión por métodos indirectos

2.1.4.1. Cálculo del nivel de nitrógeno con sensores In-situ, en tiempo real

Los sensores in-situ capaces de calcular el nitrógeno del suelo en tiempo real pueden representar una solución ideal para PA. Para ello se requiere colocar una red de sensores sobre el terreno durante el periodo de crecimiento, esto se hace para proporcionar el manejo dinámico de los fertilizantes. La ventaja del despliegue de sensores in-situ y en tiempo real es que ya no se requeriría más esfuerzo para la asimilación de datos.

A pesar de las ventajas, el hecho de colocar sensores in-situ monitoreados en tiempo real aún representa un desafío para realizar la medición directa del suelo [Shaw, 2015a]. Los desafíos a superar son:

1. Los sensores deben tener la capacidad de operar en varias condiciones ambientales y en distintos tipos de suelo. Se debe considerar que los niveles de temperatura y humedad del suelo durante el crecimiento produce resultados relevantes y desafiantes.
2. Las propiedades físicas de sensor tienen que ser robustas y duraderas para hacer frente a un periodo de entierro en un entorno agrícola. Además, deben resistir la contaminación microbiana y química de los elementos a ser sensados.
3. Los parámetros de calibración deben permanecer constante para un periodo de tiempo extendido, para que las medidas sean fiables.
4. Los sensores se deben acoplar a un sistema de registro de datos capaz de tener gran cantidad de almacenamiento. Además, estos datos necesitan ser accesibles de forma remota a través de un sistema inalámbrico.
5. Se requiere potencia para el sistema de registro de datos y para los sensores.

2.2. Aprendizaje máquina para estimación

2.2.1. El problema de regresión

Las técnicas basadas en aprendizaje máquina, buscan aprender una relación funcional entre un conjunto de variables o atributos y variables objetivo o respuestas asociadas para poder predecir la respuesta a un nuevo conjunto de atributos.

Los problemas más comunes asociados a ML pueden ser de clasificación (decisión), estimación (regresión) y agrupamiento (clustering) [Duda et al., 1973].

La estimación con ML consiste en predecir un comportamiento a partir de un modelado de datos empíricos y la tarea de la regresión consiste en encontrar una relación

entre un conjunto de observaciones ruidosas y una variable de salida real, siendo pertinente para muchas aplicaciones de ingeniería.

Supuesto que se tiene un conjunto de datos de entrenamiento $\{\mathbf{x}^{(l)}, s^{(l)}\}_{l=1}^L$, donde $\mathbf{x} \in \mathbb{R}^N$ son las observaciones y $s \in \mathbb{R}$ son las salidas deseadas, entonces, la salida de un regresor lineal toma la forma

$$\hat{s} = f(x, w) = \mathbf{w}_e^\top \mathbf{x}_e^{(l)} \quad (2.1)$$

donde $\mathbf{w}_e = \begin{bmatrix} w_0 & w_1 & \cdots & w_N \end{bmatrix}^\top$, son los pesos entrenables

y $\mathbf{x}_e = \begin{bmatrix} 1 & x_1 & x_2 & \cdots & x_N \end{bmatrix}^\top$, las observaciones extendidas.

Es posible conseguir un modelo no lineal a partir de un modelo lineal utilizando, por ejemplo expansión polinómica o alguna transformación del espacio de entrada a uno de alta dimensión utilizando funciones kernel [Rogers and Girolami, 2012].

2.2.1.1. Definición de un buen modelo

Si consideramos que $\Phi(\mathbf{X})$ es el resultado de aplicar una función de transformación a los datos, es decir:

$$\Phi(\mathbf{X}) = \begin{bmatrix} \phi(\mathbf{x}^{(1)}) \\ \phi(\mathbf{x}^{(2)}) \\ \vdots \\ \phi(\mathbf{x}^{(L)}) \end{bmatrix} \quad (2.2)$$

Entonces

$$\hat{s} = \mathbf{w}_e^\top \Phi(\mathbf{X}) \quad (2.3)$$

Para entrenar \mathbf{w}_e se minimiza una función de coste que mide el error o discrepancia entre el valor de la salida deseada y el valor de la salida estimada.

Cuando la función de coste es el promedio de los errores al cuadrado tendrá la siguiente forma

$$\mathcal{L} = \frac{1}{L} \sum_{l=1}^L \left(s^{(l)} - \mathbf{w}_e^\top \phi(\mathbf{x}^{(1)}) \right)^2 \quad (2.4)$$

y para encontrar los valores mínimos de \mathbf{w}_e se realiza la derivación a la función de coste y se iguala a cero o también se puede aplicar el método de descenso por

gradiente. Cuando se opta por derivar e igualar a cero la función el resultado es la pseudoinversa de Moore-Penrose [Duda et al., 1973] .

$$\mathbf{w}_e = \left(\Phi(\mathbf{X})^\top \Phi(\mathbf{X}) \right)^{-1} \Phi(\mathbf{X})^\top \mathbf{s} \quad (2.5)$$

Si, los pesos se entrenan por descenso por gradiente, la regla de actualización es

$$\mathbf{w}_e^{(t+1)} = \mathbf{w}_e^{(t)} - \eta \frac{\partial \mathcal{L}}{\partial \mathbf{w}_e} \quad (2.6)$$

donde t es la posición de la muestra actual y η es la tasa de aprendizaje que toma valores entre 0 y 1.

2.2.2. Generalización y sobre-ajuste

El objetivo es construir una máquina que haga predicciones con la mayor precisión, pero obtener la mejor máquina no es necesariamente aquella cuya función de coste sea la de mínimo valor. La elección de la mejor máquina está relacionada con la capacidad de generalización [Rogers and Girolami, 2012].

La capacidad de generalización de una máquina se expresa cuando el estimador es capaz de estimar de forma correcta un nuevo dato que no ha sido visto durante el entrenamiento. Cuando esto no sucede es posible que el estimador haya aprendido todos los detalles del problema y la solución este sobreajustada; por el contrario, el estimador también puede presentar un subajuste si el modelo no se ha elegido de forma adecuada y no haya alcanzado a su solución [Rogers and Girolami, 2012].

Para evitar estos problemas, es necesario validar los hiperparámetros del modelo en la etapa de entrenamiento.

2.2.2.1. Validación de parámetros libres

La validación cruzada divide el conjunto de datos de entrenamiento en bloques de igual tamaño K (también denominada *K-fold*). Cada pequeño bloque representa un conjunto de validación para un conjunto de entrenamiento formado por los otros $K-1$ bloques. Promediar los errores de los K resultados nos da el valor de validación cruzada.

A la hora de validar un parámetro libre, se elige un conjunto de posibles valores del parámetro y para cada valor se halla el valor del error de validación. Se elige como parámetro a aquel que corresponde al menor error de validación cruzada.

Existe una forma particular de validación cruzada conocida como “validación cruzada leave-one-out” (LOOCV, del inglés *leave-one-out cross validation*), es el caso extremo que se puede presentar cuando $K = L$, donde cada dato de observación es retenido y usado para probar un modelo entrenado sobre las otras $L - 1$ muestras, es decir, el modelo se entrenará L veces [Rogers and Girolami, 2012].

2.2.3. Entrenamiento con pocas muestras

En aprendizaje máquina, los conjuntos de datos son la parte principal. Este debe ser adquirido de modo que se tenga un número de muestras suficientes con instancias inherentes y que caracterizan el problema a tratar de una forma completa. Sin embargo, en muchos problemas el tomar muestras es complicado, generando un problema de pocas muestras. Un problema con pocas muestras es más sensible a las muestras atípicas, los errores de medición toman mayor importancia y pueden afectar de forma significativa al diseño del modelo final. Otra dificultad que se presenta cuando se tienen pocas muestras es a la hora de dividir los datos en entrenamiento y prueba ó cuando se realiza el procedimiento de validación cruzada para validar los parámetros libres. Si no se tiene cuidado al aplicar cualquiera de estas técnicas, se puede terminar en diseños que sobreajuste. Varios autores [Duda et al., 1973, Jiang et al., 2008] han lidiado con este tipo de problemas de diferentes formas, entre ellas podemos mencionar:

- **Revisión de los datos:** Para evitar que la variabilidad de los datos afecte la capacidad de predicción se deben detectar los valores atípicos e imputar valores perdidos de ser necesario y entender el impacto que causarán los errores de medición.
- **Uso de modelos más simples:** Utilizar modelos más simples donde el número de parámetros a validar sean pocos. Es decir, por ejemplo, en lugar de utilizar una red neuronal podemos buscar una regresión logística, o utilizar la técnica de k-vecinos más próximos en vez de regresión.
- **Regularización:** La regularización funciona bien en espacios de búsqueda pequeños, a pesar que agrega más parámetros al modelo, vale la pena, porque el resultado son modelos que dan como resultados estimaciones más robustas.
- **No hacer validación cruzada:** Si el conjunto de muestras es pequeño, lo mejor es no hacer validación cruzada para validar los parámetros libres, pero se puede optar por otros métodos como la aproximación omnisciente.
- **Intervalos de confianza:** Cuando se cuenta con pocos datos se recomienda entrenar una máquina que realice predicciones con márgenes de error, en lugar de estimaciones puntuales. Si bien es cierto que los modelos con pocos datos tienen intervalos de confianza más grandes, es mejor saber su alcance para poder tomar decisiones sobre las predicciones en lugar de no saber nada.
- **Opinión de expertos:** Un pequeño conjunto de datos, no permite probar diferentes modelos, por tanto, la opinión de un experto en el problema cuenta más. Con el apoyo de un experto, se puede conocer mejor las características del problema y ello puede apoyar a la hora de decidir qué modelo utilizar.
- **Enfoque de conjunto:** Se puede utilizar un modelo conjunto compuesto por varios modelos más simples. Los conjuntos de máquinas de aprendizaje evitan el sobreajuste sin la necesidad de aumentar el número de parámetros libres.

- Enfoque de consorcio: Generar datos en todo el espacio de observaciones. Utilizar datos con poca relación con el fin de aumentar en número de observaciones y luego restar el impacto de la no relación, puede producir modelos de mejor rendimiento.

3 Estado del Arte

Como se ha mencionado en las secciones anteriores, la presencia del nitrógeno en el suelo agrícola es de suma importancia por ser contribuyente de la clorofila y de proteínas, lo que resulta ser vital en el proceso de crecimiento de las plantas. Por ello es importante obtener una estima del nivel de nitrógeno.

No es fácil determinar de forma precisa los niveles de nitrógeno presentes en el suelo. Las raíces de las plantas pueden absorber dos formas de nitrógeno disponible en el suelo que son el nitrato y el amonio, sin embargo, generalmente se prefiere la absorción del nitrato por ser más fácilmente utilizable por la planta. Por esto, los fertilizantes contienen formas de nitrato de nitrógeno [Jeo, 1997]. Existen microorganismos que pueden cambiar de amonio a nitrato a través de un proceso llamado nitrificación.

Las formas más utilizadas para medir los niveles de amonio y nitrato (formas inorgánicas de nitrógeno) en el suelo, son mediante el uso de electrodos o por métodos colorimétricos (laboratorios) [Jeo, 1997]. El método llamado Kjeldahl ("Kel-dall") es un método estándar existente utilizado por muchos laboratorios de pruebas e investigadores que se usa para determinar los niveles de nitrógeno en el suelo.

El método utiliza ácido sulfúrico concentrado, algún tipo de oxidante (mercurio, cobre o selenio), y alta temperatura, para convertir todas las formas de nitrógeno en el suelo (orgánico e inorgánico) a amonio. Por lo tanto este método requiere el uso de algunos productos químicos muy peligrosos y aparatos muy especializados [Jeo, 1997].

En los últimos años para comprender la dinámica del nitrógeno en el suelo agrícola se han propuesto una serie de métodos o técnicas nuevas y emergentes que hagan posible un registro de medidas en el área de cultivo y en tiempo real. A continuación, daremos a conocer los más recientes y relevantes:

3.1. Muestreo del suelo basado en microdiálisis

La microdiálisis es una técnica emergente que se usa para realizar muestreos in-situ. Este método se ha utilizado recientemente para evaluar el estado de nitrógeno del suelo. En la Figura 3.1 se muestra el diagrama de una sonda de microdiálisis.

Este es un método no destructivo que se basa en la difusión pasiva de solutos desde dentro de la solución del suelo y se realiza a través de una membrana parcialmente permeable, dentro de un flujo de agua que da lugar al proceso de perfusión

[Shaw et al., 2014]. La velocidad de difusión de solutos a través de la membrana se acciona por el gradiente de concentración entre la solución del suelo que rodea la membrana y el líquido de perfusión (líquido que suministra sustancias al sistema o suelo) que está siendo bombeada continuamente a través de la sonda; este efecto crea una zona de agotamiento que rodea la sonda, que inducirá la difusión de solutos a través del suelo hacia valores menores del gradiente de concentración resultante [Shaw, 2015a].

Por lo tanto, la concentración de solutos en el dializado refleja no sólo la concentración en la solución del suelo, sino también la velocidad a la que estos solutos pueden difundirse a través del suelo [Shaw et al., 2014]. Debe comprenderse el dializado como un mecanismo químico de separación selectiva, donde una membrana separa dos soluciones que difieren en la concentración de un mismo soluto y permite el paso de ciertas soluciones y evita el de otras. La sustancia resultante del dializado se recoge sobre un período de muestreo especificado y posteriormente es analizada.

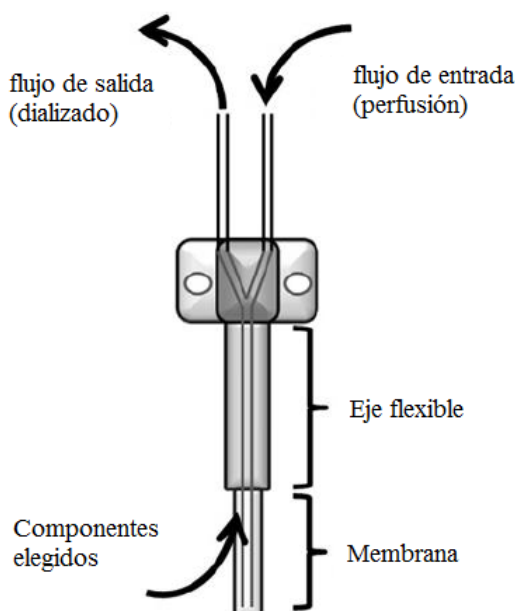


Figura 3.1: Diagrama de una sonda de microdialísis CMA 20 [Shaw, 2015a].

Para probar la tecnología de la microdialísis en el análisis de nitrógeno en el suelo y en la evaluación de las diferentes formas de nitrógeno para la nutrición de las plantas, en [Shaw et al., 2014] los autores realizaron pruebas en ocho suelos agrícolas contrastantes utilizando $0,5M K_2SO_4$ (mol de sulfato de potasio) indicaron también que en vez de esta solución es posible utilizar agua destilada obteniendo resultados similares.

Para estas pruebas las muestras del suelo se almacenaron en bolsas de polietileno permeables al gas para ser transportados inmediatamente al laboratorio, donde se

refrigeró a $<5^{\circ}\text{C}$ hasta cuando fueron requeridos para su análisis. Antes de ser usado para las pruebas, el suelo se tamizó a 8 mm para retirar la presencia de piedras grandes, raíces, vegetación y lombrices de tierra. El suelo no ha sido tamizado a 2 mm ya que en [Shaw et al., 2014] se ha demostrado que este proceso estimula la actividad microbiana dando como resultado cambios significativos de nitrógeno orgánico e inorgánico en estos suelos.

Se utilizaron dos sondas de microdiálisis con membrana de polietersulfona, la sonda se insertó al suelo mediante el uso de la aguja, antes de recoger las muestras las sondas deben permanecer por 20 minutos en el interior; los flujos difusivos de los AA (aminoácidos), amonio y nitrato se determinaron mediante el cálculo de la cantidad de cada forma de nitrógeno que es difundido a través de la membrana de la sonda de microdiálisis durante cada periodo de muestreo.

Por los resultados de concentración de solutos en el dializado se ha argumentado en [Shaw et al., 2014] que la microdiálisis es el mejor método para evaluar el nitrógeno del suelo ya que es el más confiable en la entrega de informes sobre la disponibilidad de nitrógeno y la importancia de las diferentes formas de nitrógeno para la nutrición de las plantas. Además, las sondas de microdiálisis, por ser de tamaño relativamente pequeño y de corto período de muestreo, ofrecen una muy buena resolución espacial y temporal.

Sin embargo, aún presenta inconvenientes por las complejidades de calibración, es costosa de realizar, las sondas sólo muestrean pequeños volúmenes de suelo, y además el equilibrio entre las concentraciones de soluto de la solución de diálisis y la solución del suelo nunca se logra por el **contenido de humedad, la temperatura, tamaño de partícula, y la composición** [Shaw et al., 2014].

A pesar de estos inconvenientes, la simple medición de la velocidad de difusión de solutos a través de la membrana de la sonda con el tiempo puede dar un método útil para evaluar el nitrógeno disponible en el suelo [Shaw et al., 2014].

3.2. Método de electrodos de ión selectivos

3.2.1. Electrodos selectivos de iones de nitrato — in-situ

El método a través de electrodos de ión selectivos (ISEs, del inglés *ion-selective electrodes*) responden a la concentración sensada o detectada por las plantas y microbios. El uso de ISEs para determinar la actividad o concentración de nitrato en el suelo agrícola han sido usadas para las pruebas rápidas y para los muestreos tomados en marcha. Para determinar el nitrato la mayoría de los ISE cuentan con una membrana selectiva de matriz polímero PVC que contiene un cóctel químico de detección inmiscible con agua [Shaw, 2015a, Shaw et al., 2013].

En la Figura 3.2 se muestra el sistema de trabajo del ISE para realizar la medida de la diferencia de potencial entre un electrodo de referencia y el electrodo que contiene

una membrana la que es seleccionada para el ión de interés [Shaw, 2015a].

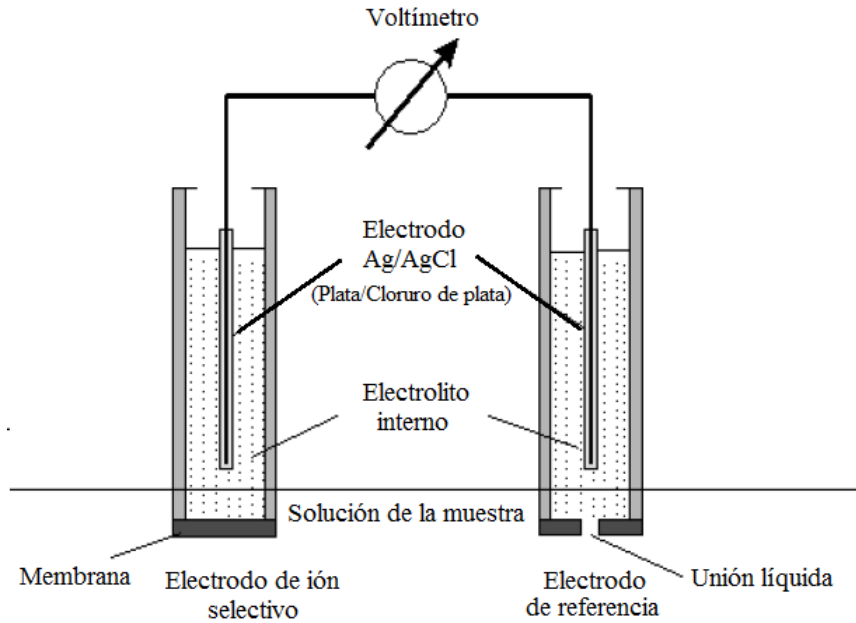


Figura 3.2: Representación esquemática de un sistema ISE, mostrando que el ISE y el electrodo de referencia son enlazados por un voltímetro [Shaw, 2015a].

La diferencia de potencial está linealmente relacionada con el logaritmo de la actividad del ión seleccionado y se describe por la ecuación de Nernst (Ecuación 3.1) [Shaw, 2015a].

$$E = G + \left(\frac{2,303RT}{z_i F} \right) \log a_i \quad (3.1)$$

Donde E es la diferencia de potencial, z_i y a_i son la carga y la actividad del ión de interés, G es una constante que incluye todas las contribuciones potenciales de muestras independientes (influenciado por el diseño del ISE), R es la constante del gas, T es la temperatura en Kelvin y F ($Cmol^{-1}$) es la constante de Faraday [Shaw, 2015a].

Los ISEs para nitrato se han utilizado en experimentos con trigo durante el invierno según [Shaw, 2015a] para determinar la concentración de nitrato en el suelo. Este método implica que se agregue 1mL de agua destilada en un agujero in-situ, esto se debe dejar por 10 minutos para que se pueda equilibrar con el suelo y después el ISE se usa para medir directamente los niveles de nitrato. A veces, cuando el suelo está seco y agrietado, una muestra del suelo es removida y se mezcla con agua para formar una pasta, la que después es medida con ISE; aunque este último método usado no es estrictamente una medida del suelo directa (DSM, del inglés *Direct*

Soil Measurement), en [Shaw, 2015a] se mostró el potencial de los ISEs para ser utilizados de esta manera.

El experimento realizado por [Shaw et al., 2013] involucró tres tipos de suelo, uno era de textura arcillosa, usada para la producción de papa en la campaña anterior, el segundo suelo recibió fertilizantes regulares, y el tercero fue un suelo arenoso con pastoreo extensivo. El suelo se almacenó en bolsas de polietileno permeables al gas para el transporte inmediato al laboratorio, se refrigeró a $<5^{\circ}\text{C}$ y antes de su uso, el suelo se tamizó a 8 mm para eliminar grandes piedras, raíces, la vegetación y las lombrices de tierra y luego se mezcla a fondo.

Los estudios en [Shaw et al., 2013] demostraron que la extracción de nitrato con cloruro de potasio (KCl), o con agua (H_2O), tienen resultados muy similares, por lo que hacerlo con H_2O resulta totalmente aceptable ya que no altera los resultados.

Una de las ventajas de los ISEs es que se pueden utilizar en suelo húmedo sin necesidad de filtración o centrifugación. En [Shaw, 2015a] se considera que este método es económico, relativamente preciso o exacto y que ofrece un tiempo de respuesta corto, el cual es menor a 60 segundos, todo dentro de un paquete portátil, por lo que resulta muy útil en el seguimiento realizado en campo.

El uso de ISEs resulta ser muy efectivos, pero uno de sus inconvenientes es la durabilidad del electrodo puesto a prueba en el suelo, para ello es necesario el diseño de una carcasa del sensor más resistente que pueda mejorar esta falta de durabilidad junto con los cambios en la química del sensor [Shaw et al., 2013].

A pesar de sus ventajas, este método presenta problemas a largo plazo debido a la desviación de calibración produciendo en la ecuación de Nernst (Ecuación 3.1) un cambio en la pendiente, además de un incremento de su sensibilidad a la temperatura, al pH del suelo y a la humedad, además de su poca durabilidad por la contaminación de sus membranas. Sin embargo a corto plazo estas medidas son fácilmente controladas [Shaw, 2015a].

3.2.2. Electrodo selectivo de iones in-situ en tiempo real

Estos nuevos electrodos selectivos de ión de nitrato dan nuevas oportunidades de medidas in-situ y en tiempo-real de la dinámica del nitrógeno del suelo, proporcionando un adelanto para la ciencia del suelo analítico y el monitoreo ambiental.

La dinámica de los nutrientes en el suelo resulta algo limitado por la incapacidad de observar y por los procesos de medida en tiempo real, sin destruir o perturbar el sistema que está siendo evaluado. En la agricultura la información en tiempo real e in-situ proporciona información sobre la dinámica espacial y temporal de los nutrientes disponibles para las plantas tales como el nitrógeno [Shaw, 2014].

Los electrodos de ión selectivo se han utilizado para la medición de nitrato en pequeñas áreas de suelo agrícola, pero en el [Shaw, 2014] se demuestra el uso de los

ISE para realizar medidas del nitrato en un suelo agrícola monitoreadas en tiempo real. Los resultados son similares a los análisis de laboratorio convencional.

Los ISE son capaces de medir directamente una amplia gama de analitos (sustancia que puede ser un ión, un elemento o un compuesto, que posee muestras que deseamos analizar), pero presentan una serie de problemas cuando se usan en el análisis de suelos pues en el proceso los electrodos son ensuciados, lo que conlleva a cierta inestabilidad [Shaw, 2014].

Las pruebas que se realizaron fueron la extracción de nitrato usando ISEs. Para estas pruebas en [Shaw, 2014] se necesitó adicionar 1 ml de agua destilada a un agujero in-situ (mantenido por tubos de polietileno) en el suelo. Luego de 10 minutos (tiempo que permite que el ISE se equilibre con el suelo) se midió la concentración de nitrato [Shaw, 2014]. Además se probó con ISEs que contenían PVC de membrana base, pero sus resultados fueron de baja precisión; un nuevo tipo de ISE que utiliza una matriz de Urushi (un tipo de látex) de membrana, mostró un alto nivel de acuerdo con la determinación de laboratorio basado en colorimetría convencional.

Para evaluar las características químicas del suelo, se recogieron muestras de 4 bloques (30×30 cm²) a una profundidad de 0-10 cm. El suelo se colocó en bolsas de polietileno permeables a los gases y se transportó al laboratorio en una caja refrigerada.

Durante los dos periodos de monitoreo sólo 1 de los 17 ISEs de nitrato falló. El único fallo ocurrió durante la inserción del electrodo en el suelo, lo que causó que la membrana se deslice hacia arriba en el cilindro del electrodo. Demostrándose que los ISE de nitrato son suficientemente robustos para la supervisión in-situ.

Se debe aclarar que los ISEs de nitrato detectan la actividad y no la concentración de los iones de nitrato en la solución del suelo, dicha actividad refleja una medida de concentración de nitrato más baja con respecto a la concentración absoluta (aproximadamente 8 % más baja) [Shaw, 2014].

Los resultados en [Shaw, 2014] muestran que la movilidad de nitrato en el suelo, se incrementará si este se vuelve más húmedo, lo que puede reducir su heterogeneidad. Además, observó que las altas temperaturas en el día involucran una variación en los resultados de los ISE de nitrato en el experimento de campo que son mucho más grandes que en el experimento de laboratorio, esto puede ser porque la temperatura del suelo a ser transportada es mucho más baja. Los resultados de los ISEs de nitrato aumentaron dramáticamente a medida que la temperatura del registrador de datos excede 25 °C de una manera que era consistente con una relación exponencial entre la temperatura y la salida ISE. La temperatura puede estar relacionada con la concentración de nitrato del suelo. Como tal, es probable que cada ISEs de nitrato tenga una respuesta ligeramente diferente ante la temperatura. Idealmente, la temperatura responde a cada ISE de nitrato que sería caracterizada individualmente, lo que la hace muy lenta y sin certeza de que esto mejoraría la precisión de la monitorización in-situ. Para futuros estudios se recomienda que el registrador de datos del ISEs de

nitrito se almacene en un recipiente aislado térmicamente, de la luz solar directa, a fin de mantener una temperatura más estable [Shaw, 2014].

3.3. Red de sensores in-situ

Para una mejora de la eficiencia in-situ en [Shaw, 2015b] se desplegó sensores y se usó un método de muestreo anidado geo-estadístico para caracterizar la variabilidad espacial de los aminoácidos, amonio y nitrato.

Para lograr una agricultura de precisión (PA del inglés *Precision Agriculture*) y reducir las pérdidas de nitrógeno, en [Shaw, 2015b] se usa sensores in-situ capaces de monitorizar el nitrógeno mineral del suelo en tiempo real. Este enfoque puede permitir dejar de aplicar los fertilizantes de nitrógeno de manera predeterminada para empezar a utilizar un sistema más dinámico que responde en tiempo real a los cambios en las condiciones de cultivo [Shaw, 2015b]. Los datos proporcionados por los sensores in-situ serán de mayor resolución espacial, por lo que no es necesario la toma de muestra de suelo repetida a lo largo del año, reduciendo el coste económico para el agricultor. Además, esto puede permitir la sincronización de la aplicación de fertilizantes, lo que reduce el riesgo de penalizaciones de rendimiento causados por deficiencias de nutrición de nitrógeno, y reducir el riesgo de transferencias de nitrógeno al agua y el aire como resultado de aplicaciones excesivas de fertilizantes de nitrógeno [Shaw, 2015b].

La optimización de la configuración espacial de una red de sensores asegura que la estimación precisa de la media a través de una zona o campo de gestión puede hacerse mientras que a la vez se minimiza los costes económicos. Por lo tanto, la evaluación de la variación espacial del nitrógeno mineral del suelo a través de una gama de escalas permite determinar una configuración óptima antes de la implementación de la red de sensores. El despliegue de sensores depende de la variabilidad de las propiedades a diferentes escalas dentro de campo [Shaw, 2015b], por ejemplo, en un campo que tiene poca variación a gran escala puede ser servido por una colección de sensores conectados a un concentrador de detección, mientras que un campo con una mayor variación a escalas más grandes requeriría sensores que se encuentra en múltiples áreas del campo.

Se realizaron dos campañas de muestreo durante el periodo de crecimiento, las que fueron tomadas antes y después de la aplicación de fertilizantes nitrogenados, el suelo fue fertilizado dos veces en un año a razón de $N: P: K$ 50:10:10 y 60: 4: 0 $kg ha^{-1}$, respectivamente.

Para el análisis de las características del suelo en [Shaw, 2015b] se recogieron muestras de 4 bloques de $30 \times 30 cm$ a una profundidad de $0-10 cm$, el suelo se colocó en bolsas de polietileno permeables a los gases y se transportó al laboratorio en una caja refrigerada.

3.3.1. Muestreo anidado para la medida de variabilidad espacial

Los sensores fueron desplegados alrededor del único registrador de datos, teniendo una distancia máxima de $2m$ entre cada sensor. Para ello se utilizaron redes de sensores anidados cuyos nodos están separados por distancias de $1cm$, $10cm$, $50cm$ y $2m$. La Figura 3.3 muestra el diseño de la red, con 12 nodos por estación principal. Las muestras de suelo fueron tomadas a una profundidad de $5-10\text{ cm}$ por ser la zona de enraizamiento, luego se colocaron en bolsas de plástico permeables al gas, y se almacenaron en una caja refrigerada para ser transportadas inmediatamente al laboratorio donde fueron refrigeradas a 4°C . Las distancias entre los puntos de muestreo se fijaron pero fueron aleatorizados los ángulos (con la excepción de los vectores de $2m$).

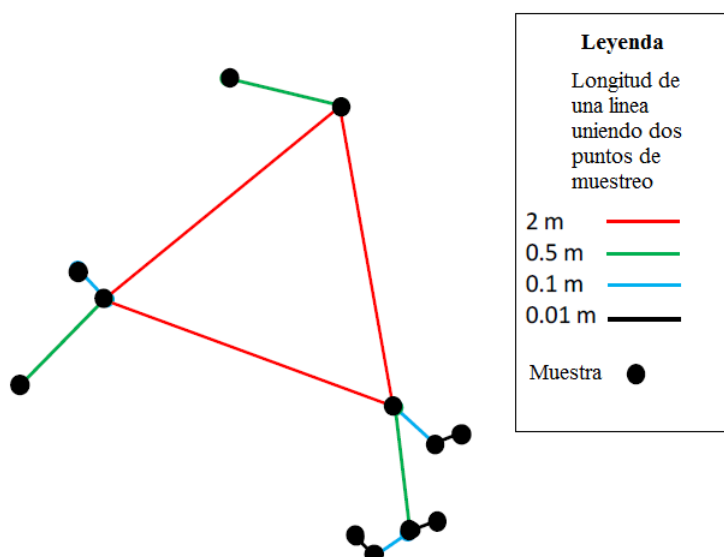


Figura 3.3: Diseño de muestreo optimizado de una estación principal [Shaw, 2015b].

En [Shaw, 2015b] se muestra que la escala de $50cm$ fue elegida como el muestreo anidado inicial por abarcar la mayor parte de la varianza para todas las formas de nitrógeno [Shaw, 2015b].

En la Figura 3.4 podemos observar el comportamiento de los niveles de nitrógeno con respecto a la profundidad de las medidas realizadas, así a mayor profundidad la concentración de los nutrientes va disminuyendo.

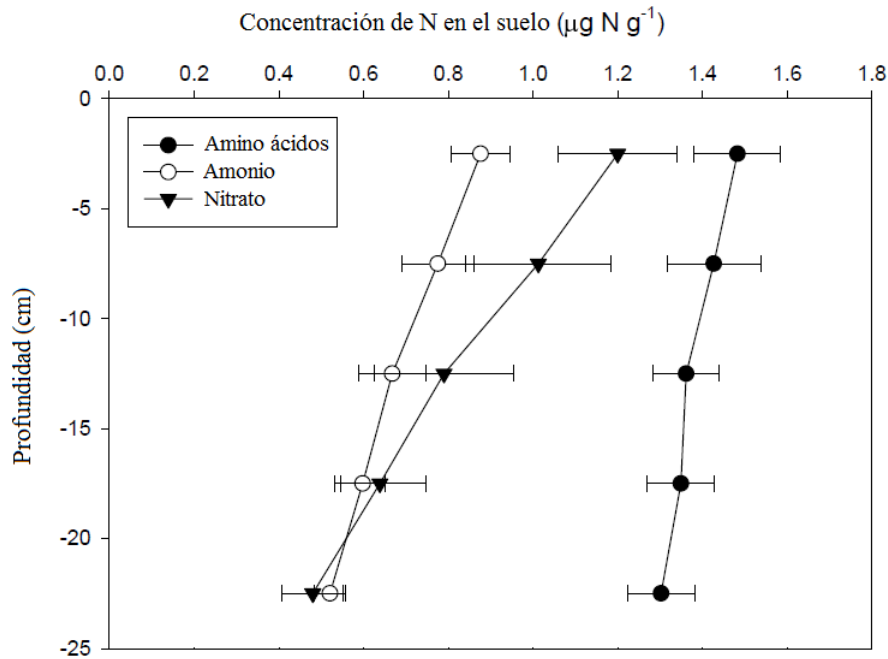


Figura 3.4: Concentración de nitrógeno del suelo para cada incremento de profundidad de 5 cm [Shaw, 2015b].

3.3.2. Optimización de una red de sensores dentro de campo para el control de nitrógeno soluble en el suelo

Para determinar la dinámica de los niveles de nitrógeno en el suelo agrícola, los autores de [Shaw, 2015b] realizaron cálculos para registradores entre 1 y 10, siendo 15 el número máximo de puertos de sensores en el registrador de datos. Sus experimentos realizados demuestran que al aumentar del número de sensores por registrador, y aumentar el número de registradores, reduce la anchura del intervalo de confianza. Pero la reducción de esta anchura sustancialmente requeriría costos monetarios sustanciales.

Si se requiere una amplitud del intervalo de confianza del 95 % [Shaw, 2015b] realizó experimentos con tres opciones: uso de 3 registradores con 11 sensores por registrador, 4 registradores con 5 sensores por registrador y 5 registradores con 4 sensores por cada registrador. Los costos de estas opciones involucran mayor costo a medida que se utilizan más registradores. Esta medida de la calidad no se puede lograr con sólo una o dos registradores. Por lo tanto, la elección racional de la configuración de la matriz es de 3 registradores con 11 sensores en cada uno.

Además en [Shaw, 2015b] se ha demostrado que es posible que el despliegue de electrodos selectivos de Nitrato se usen con éxito para los flujos de monitoreo y para

medidas rápidas de nitrato, aunque aún se requieren de mejoras en la precisión y en robustez de la membrana de detección.

3.4. Otros sensores

Existe una variedad de sensores amperométricos/voltamétricos, y biosensores, que pueden ser usados para la determinación de nitrato en solución [Shaw, 2015a]. Sin embargo, aún se necesita llevar a cabo más experimentos e investigaciones para que se asigne un adecuado análisis de suelos. En general su funcionamiento permite la reducción de nitrato en la superficie del electrodo, creando una respuesta electroquímica medible, la que se relaciona con la concentración de nitrato en la solución.

3.4.1. Nano sensores electroquímicos

El avance tecnológico ha dado lugar al desarrollo de una variedad de nanos sensores electroquímicos, y es probable que se alcancen grandes avances en los próximos años. Este es un sensor electroquímico fabricado por electrodeposición de nano partículas de plata sobre polipirrol pre-sintetizado (PPy), para formar una matriz compuesta PPy/Ag sobre un electrodo de carbono vítreo (vítreo es un material semejante al vidrio que posee alguna de sus características). Este sensor mostró una muy buena sensibilidad, selectividad y estabilidad [Shaw, 2015a].

3.4.2. Nano sensores de electrodos de alambre

Los nanos sensores de electrodos de alambre han mostrado una ventaja por su tiempo de respuesta corta, mayor sensibilidad y menores volúmenes de muestra, pero estos sensores aún no se han utilizado para la determinación del nitrato y amonio del suelo [Shaw, 2015a]. El inconveniente con los nanos sensores es que debido a su pequeña escala espacial con la que trabajan es que se requiere de mayores conjuntos de sensores.

3.4.3. Biosensores

Los biosensores permiten medir el nitrato en una variedad de ambientes, incluyendo suelos tipo lodo, alimentos, aguas residuales y agua de mar. El inconveniente es que su tiempo de vida es muy corto y presentan carencia de robustez [Shaw, 2015a].

3.4.4. Tecnología de laboratorio sobre un chip

El avance tecnológico ha permitido el desarrollo de la miniaturización de las técnicas de análisis químico, que normalmente se llevan a cabo en el laboratorio. Esto permite

el monitoreo ambiental in-situ. En [Shaw, 2015a] se han desarrollado pruebas en aguas naturales, pero aún no se han utilizado en pruebas de suelo. Otro inconveniente es que este tipo de análisis resulta muy caro por lo que no es adecuado como método de pruebas regulares y frecuentes.

4 Propuesta de Solución

Los métodos descritos en la Subsección 2.1.3 y Subsección 2.1.4 para medir los niveles de nitrógeno en suelo agrícola presentan dificultades a la hora del trabajo en campo. Por un lado, requieren la toma de muestras y el envío a laboratorio y por otro, el despliegue en campo de instrumentación que requiere de personal cualificado. En los últimos años se ha notado un gran avance en el uso de la tecnología para el riego agrícola. En la mayoría de casos se tiene información en tiempo real del clima y las condiciones ambientales del suelo. Entre las variables que con mayor frecuencia se miden en el suelo están la tensión del suelo como indicador de humedad y el volumen de agua. Otra de las variables es la temperatura y la conductividad eléctrica. Como ya se mencionó en la Subsección 2.1.1, las pérdidas de nitrógeno están relacionadas con la humedad del suelo. Por otro lado, un indicador de su presencia es la conductividad eléctrica dado que este cambia con la presencia de sales en el suelo. En base a este análisis, una forma de tener información de los niveles de nitrógeno en suelo es considerar como entradas de un estimador máquina las condiciones ambientales del suelo y con ello estimar los niveles de nitrógeno en suelo agrícola. Para saber qué tipo de estimador utilizar, primero vamos a generar una base de datos etiquetadas de las variables que son relevantes para estimar los niveles de nitrógeno en el suelo. Para ello, debemos identificar aquellas variables relevantes que sean medibles en tiempo real. Según el comportamiento de los niveles de nitrógeno en el suelo (etiquetas) se procederá a diseñar el modelo máquina que puede ser lineal o no lineal. Para validar el modelo propuesto, utilizaremos datos reales adquiridos en un campo de cultivo en etapa de fertilización. Las prestaciones del modelo se evaluarán utilizando una función de coste que mide la discrepancia entre el valor deseado (etiqueta) y el valor estimado, cuyos resultados se presentarán como el promedio de los errores y su respectiva desviación típica. Para validar los parámetros libres del modelo propuesto, utilizaremos la técnica de validación cruzada. Lo antes expuesto se ha desarrollado en diferentes actividades que se detallan en función a cada objetivo específico.

4.1. Toma de muestras de las condiciones ambientales del suelo

Se escogió dos tipos de plantaciones, una de alcachofa y otra de paprika. Para cada caso, se colocaron sensores de conductividad eléctrica, temperatura y humedad en un mismo punto. El sistema de riego es por goteo y trabaja de manera automática en base a una plataforma inteligente que lee las condiciones ambientales del terreno

mediante una red de sensores (nodos terminales) y en función de ello, envía órdenes para la apertura o cierre de las válvulas de riego (nodos ruteadores). Todo el sistema es coordinado por un servidor (nodo coordinador) que posee un algoritmo de control predictivo. El esquema de la plataforma inteligente se muestra en la Figura 4.1.

Se ha ubicado en un punto de manera aleatoria los sensores de temperatura, conductividad eléctrica y humedad, integrados en un mismo dispositivo de la marca DECAGON modelo 5TE, a una profundidad de 20 cm y a una distancia de 5 cm de la línea de riego [Decagon, 2016].

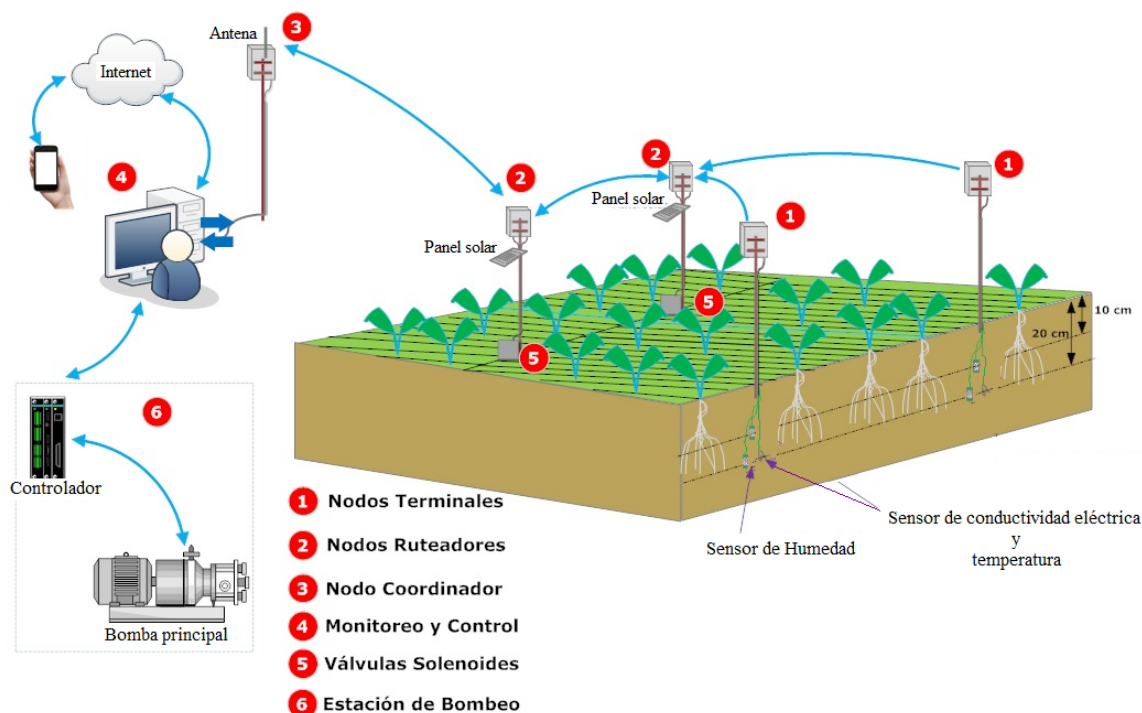


Figura 4.1: Despliegue de sensores en la plataforma inteligente para el sistema de riego.

Los sensores están conectados a un nodo terminal, el cual se conecta de manera inalámbrica a nodos ruteadores y estos a un nodo coordinador conectado a un ordenador localizado en su centro de operaciones donde se almacenan todas las medidas adquiridas, formando así una base de datos en tiempo real.

4.2. Toma de muestras de los niveles de nitrógeno en suelo

Las muestras tomadas deben realizarse siempre en la misma época o con los mismos cultivos [Germán, 2013]. Por ello para el experimento se decidió observar el comportamiento de dos terrenos diferentes, uno cultivado con alcachofa y otro cultivado

con paprika. El terreno cultivado con alcachofa se encontraba en su segunda semana de crecimiento por lo que las cantidades de nitrógeno aplicado era muy bajo, en cambio en el cultivo de paprika las plantaciones se encontraban en la última semana de crecimiento por lo que era mayor la cantidad de nitrógeno aplicado.

Antes de tomar las muestras, se ha desechado los primeros *5cm* según se recomienda en [Germán, 2013]. Luego se introduce la pala a una profundidad de *20cm* y se cava haciendo un hoyo en forma de "V". De este hoyo, se extrajo entre *500g* y *1kg* de tierra utilizando una cuchara. Todas las muestras fueron almacenadas de forma directa en una bolsa de plástico con cierre hermético y enviadas al laboratorio para su análisis.

Las muestras fueron tomadas cada media hora en cada uno de los terrenos (alcachofa y paprika). La primera muestra fue obtenida antes de regar la planta con fertilizante, las siguientes se sacaron después de ser regadas con fertilizantes de nitrógeno.

En la Figura 4.2 se observa las muestras embolsadas y etiquetadas para su análisis en el laboratorio, que fueron 15 muestras para la paprika y 14 muestras para la alcachofa.



Figura 4.2: Muestras de tierra agrícola etiquetadas para su análisis de nitrógeno en el laboratorio.

4.3. Análisis de las muestras

Para encontrar la relación entre los parámetros de entrada (temperatura, conductividad eléctrica, humedad) y el comportamiento de los niveles de nitrógeno en el suelo agrícola se graficará sus valores reales medidos con respecto al tiempo, tanto para el cultivo de paprika como para la alcachofa.

Paprika

Las plantaciones de paprika se realizaron en un terreno franco arenoso, este suelo tiene una textura equilibrada; es decir, relativamente suelta, por lo que es apto para este tipo de cultivo.

En la Figura 4.3a podemos observar el comportamiento de la temperatura, conductividad eléctrica, humedad y niveles de nitrato (nitrógeno), la toma de muestras de suelo para su análisis se realizaron aproximadamente cada media hora, desde las 9:00 hasta las 16:05 horas, dando lugar a un número total de 15 muestras. Asignando 10 muestras para el entrenamiento del modelo máquina y 5 muestras para hacer pruebas.

El día de la obtención de muestras el terreno fue regado alrededor de las 9:30 horas, por lo que se puede observar en el gráfico de humedad de la Figura 4.3a que el suelo hasta antes de las 9:40 horas tiene un volumen de contenido de agua bajo pero diferente de cero ya que el suelo aún preserva humedad del día anterior regado; desde las 9:40 horas hacia adelante el suelo aumenta sus niveles de humedad a causa del riego, el que dura aproximadamente hasta las 10 horas. A partir de este punto, los niveles van descendiendo por la ausencia de agua suministrada; es decir, que el suelo se empieza a secar y el principal causante de este fenómeno es la temperatura que va incrementando a medida que pasan las horas, haciendo que el suelo se caliente y se seque.

También podemos observar en la figura que los niveles de conductividad eléctrica ascienden cuando se tiene la humedad máxima, esto se debe a que este cultivo está siendo regado con nutrientes de un gran contenido de nitrógeno. Por ello como el suelo está húmedo y rico en minerales la conductividad aumenta.

El nivel del nitrógeno desciende al aumentar la humedad, esto se debe al proceso de lixiviación que hace que por acción del riego los nitratos (nitrógeno nítrico) se muevan fácilmente hacia abajo. Otro motivo es por la absorción del nitrógeno por parte de las plantas.

Alcachofa

Las plantaciones de alcachofa se realizaron en un terreno arenoso, cuya característica es que cuando es humedecido no se engancha, es decir que el suelo no retiene el agua, haciendo que esta se filtre a capas más profundas. Además, el tipo de riego para la paprika es diferente al de la alcachofa, ya que este último requiere de mayor cantidad de agua para su crecimiento, por ello debe ser regado con mayor frecuencia y periodo de tiempo.

Como se puede observar en la Figura 4.3b el comportamiento de la temperatura, conductividad eléctrica, humedad y niveles de nitrato (nitrógeno) son distintos al de paprika. Para explicarlo primero empezaremos diciendo que la toma de muestras

4.4 Diseño de un modelo máquina para la estimación de los niveles de nitrógeno en suelo agrícola

de suelo para su análisis en laboratorio se realizaron aproximadamente cada media hora, desde las 9:27 hasta las 15:59 horas, dando lugar a un número total de 14 muestras. Asignando 9 muestras para el entrenamiento del modelo máquina y 5 muestras hacer pruebas.

El día de la obtención de muestras el terreno fue regado alrededor de las 11:00 horas, por lo que se puede observar en el gráfico de humedad de la Figura 4.3b que el comportamiento del suelo hasta antes de las 11:00 horas tiene un volumen de contenido de agua bajo pero diferente de cero por motivos explicados en la sección anterior; desde las 11:00 horas hacia adelante el suelo aumenta sus niveles de humedad a causa del riego, el que dura aproximadamente hasta las 12:15 horas, los niveles de humedad son mayores que los de paprika ya que como se mencionó el cultivo de alcachofa requiere mayor cantidad de agua. En este caso la temperatura no le afecta en gran medida a la humedad para su descenso, pues al estar la tierra muy húmeda, esta tardará en secar.

También podemos observar que los niveles de conductividad eléctrica descienden a medida que la humedad incrementa, como dijimos la conductividad eléctrica aumenta cuando el suelo es regado con nutrientes, pero en este caso las plantas de alcachofa tenían poco tiempo de ser sembradas (pocas semanas) por lo que el riego se realiza con muy pocas cantidades de nutrientes, además este día se regó con agua libre de nutrientes, para hacer un lavado del suelo, como ellos lo llaman. Por este motivo es que la conductividad eléctrica baja a medida que se riega ya que se va lavando los minerales (lixiviación).

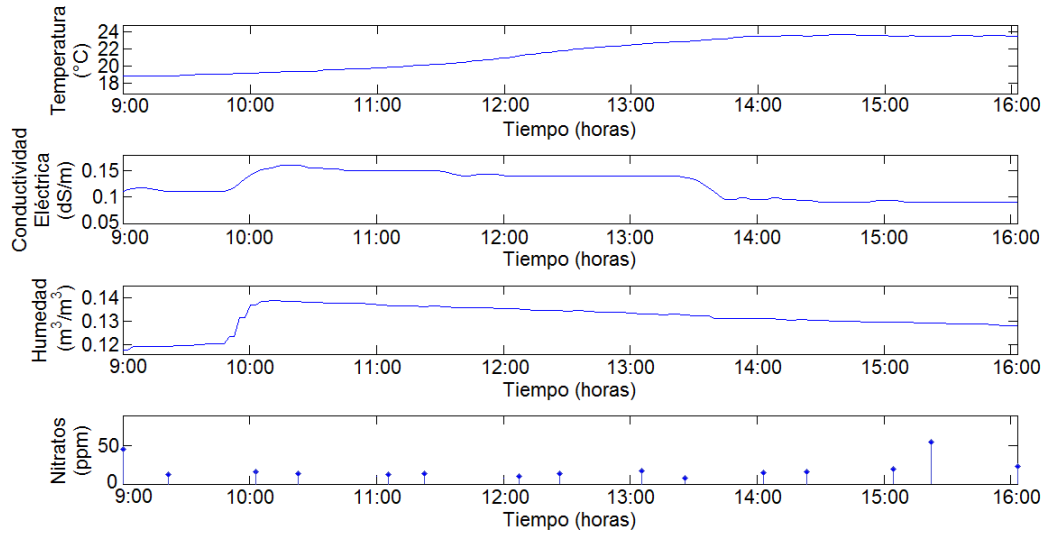
Al observar el nivel de nitrógeno en la alcachofa nos damos cuenta que estos son mayores que los de paprika, esto se explica por la cantidad de nitrógeno que dejó la cosecha anterior, además que el terreno antes de ser sembrado fue abonado, y al ser las plantas aún muy pequeñas estas no absorben grandes cantidades de nutrientes.

A la vista de estos análisis, podemos ver una alta correlación entre los niveles de Nitrógeno y las medidas entregadas por los sensores de conductividad eléctrica, temperatura y humedad. Por tanto, es posible diseñar un modelo máquina para predecir el comportamiento del nitrógeno a partir de las medidas de los niveles de humedad, temperatura y conductividad eléctrica, dado que estos valores se pueden obtener de manera fácil y en tiempo real.

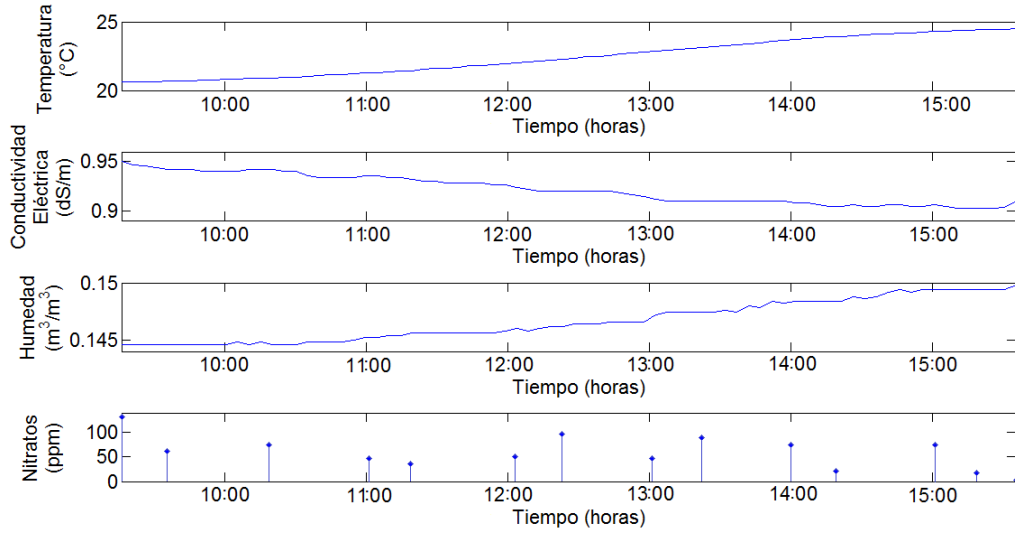
Como se puede ver, el comportamiento del nitrógeno es de forma no lineal. Por ello elegimos un modelo máquina no lineal.

4.4. Diseño de un modelo máquina para la estimación de los niveles de nitrógeno en suelo agrícola

Para estimar el nivel de nitrógeno en el suelo agrícola utilizaremos un modelo no lineal con Kernels Gaussianos, es decir:



(a)



(b)

Figura 4.3: Condiciones ambientales de temperatura, conductividad eléctrica, humedad y nitratos medidos en los terrenos cultivado con: (a) paprika y (b) alcachofa.

$$\mathbf{F}(\mathbf{x}) = \mathbf{w}_e^\top \Phi(\mathbf{x}) \quad (4.1)$$

donde \mathbf{w}_e es el vector de pesos del estimador y $\Phi(\mathbf{x}) = [\Phi_0, \Phi_1(x), \Phi_2(x), \dots, \Phi_M(x)]^\top$ el vector de salida de funciones gaussianas que tienen la forma:

$$\Phi_m = \begin{cases} 1, & m = 0 \\ \exp\left(-\frac{\|\mathbf{x} - \mathbf{C}_m\|^2}{2\sigma_m^2}\right), & 1 \leq m \leq M \end{cases} \quad (4.2)$$

$\{\mathbf{C}_m\}$ son los centros de las funciones gaussianas y $\{\sigma_m\}$ sus correspondientes dispersiones.

Generalmente, el entrenamiento que se realiza en este tipo de máquinas es en dos etapas: en la primera etapa se aprenden las funciones base y sus respectivos parámetros. En una segunda etapa, los pesos de salida del estimador [Mayhua-López, 2012].

Para encontrar los valores óptimos del vector de pesos, vamos a utilizar el criterio de mínimos cuadrados, cuya función de coste a minimizar es:

$$\mathcal{L} = \frac{1}{L} \sum_{l=1}^L \left[s^{(l)} - F(\mathbf{x}^{(l)}) \right]^2 \quad (4.3)$$

Por lo tanto

$$\mathbf{w}_e^* = \arg \min_{w_e} \{\mathcal{L}\} \quad (4.4)$$

Es decir

$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}_e} = 0 \quad (4.5)$$

$$\frac{1}{L} \sum_{l=1}^L 2 \left(s^{(l)} - \mathbf{w}_e^\top \Phi(\mathbf{x}^{(l)}) \right) \left(-\Phi(\mathbf{x}^{(l)}) \right) = 0 \quad (4.6)$$

Procediendo a simplificar tenemos que

$$\mathbf{w}_e = \left(\Phi(\mathbf{x})^\top \Phi(\mathbf{x}) \right)^{-1} \Phi(\mathbf{x})^\top \mathbf{S} \quad (4.7)$$

De modo que

$$F(x) = w_0 + w_1 \Phi_1(x) + w_2 \Phi_2(x) + \dots + w_M \Phi_M(x) \quad (4.8)$$

Otra forma de entrenar los pesos es por algún algoritmo de búsqueda como el descenso por gradiente. Dado que la función de coste que se utiliza es del error cuadrático medio, es decir una función convexa, el valor mínimo o máximo está garantizado, por tanto, también la convergencia. Esto es, cuando utilizamos la Ecuación 4.4, de

forma directa se halla el valor mínimo. Si se utiliza descenso por gradiente, con elegir una tasa de aprendizaje adecuada junto con el número de épocas necesarias, la convergencia está asegurada y se deben tener resultados similares.

Para aprender las funciones base y sus respectivos parámetros, primero fijamos los centroides y luego los parámetros de dispersión de las funciones gaussianas.

Selección de centroides

Para la selección de los centroides y evitar incluir centroides redundantes entre sí por estar localizados en la misma región del espacio de observaciones, se aplica un procedimiento similar al algoritmo de clustering APC-III (Adaptive Patern Classifier III) [Young-Sup and Sung-Yang, 1996]. Se elige como primer centroide la muestra que presenta mayor distancia hacia todas las muestras. Una vez seleccionado el primer centroide, se eliminan del conjunto de muestras aquellas que estén a una distancia menor que h del centroide que se acaba de seleccionar. Con el conjunto de muestras restantes, se repite este proceso hasta que no queden más disponibles.

En valor de h se calcula como el promedio de la distancia mínima entre las muestras de entrenamiento multiplicado por un factor R ,

$$h = R \frac{1}{L} \sum_{l=1}^L \left\| \mathbf{x}^{(l)} - \mathbf{x}_{NN}^{(l)} \right\|_2 \quad (4.9)$$

donde $\mathbf{x}_{NN}^{(l)}$ indica el vecino más cercano a $\mathbf{x}^{(l)}$.

Selección de dispersión de las funciones gaussianas

Para la selección de los parámetros de dispersión de las funciones gaussianas se seleccionarán valores proporcionales al promedio de la distancia euclídea entre C_m y todas las muestras que se encuentran más cercanas a él que del resto de los demás centroides, siendo la ecuación

$$\sigma_m = r \frac{1}{\#C_m} \sum_{\mathbf{x}^{(l)} \in C_m} \left\| \mathbf{x}^{(l)} - \mathbf{c}_m \right\|_2 \quad (4.10)$$

donde C_m representa el conjunto de muestras que pertenecen al m -ésimo centroide, $\#$ la cardinalidad y r el factor de escala.

5 Experimentos y Resultados

Para validar el modelo máquina propuesto para la estimación del nivel de nitrógeno en el suelo, hemos realizado experimentos sobre una base de datos reales. La base de datos se ha generado en una plantación de paprika y alcachofa en el fundo Santa Gabriela localizado en la Irrigación Santa Rita de Sigüas perteneciente a la ciudad de Arequipa, tal como se describe en la Sección 4.1 y Sección 4.2.

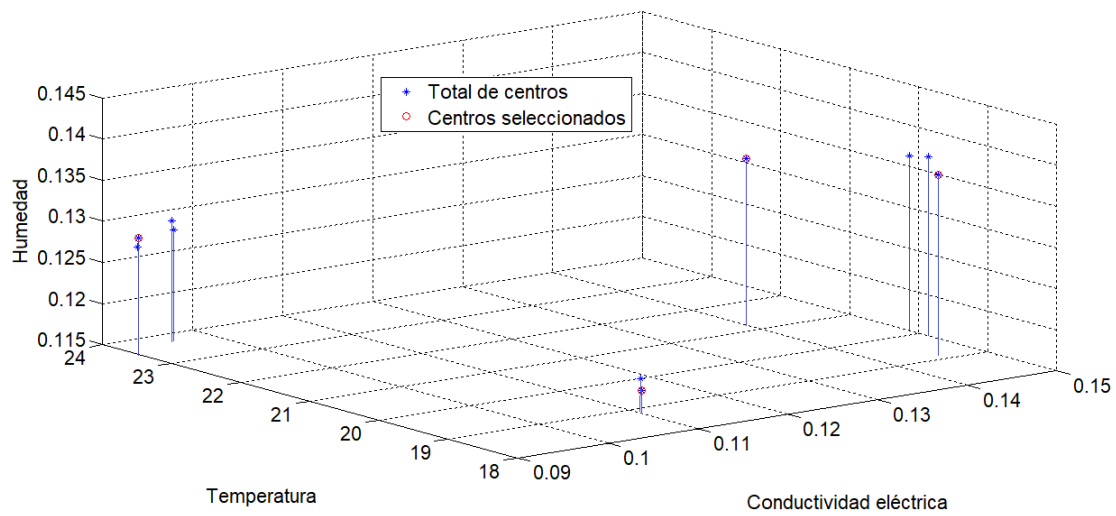
5.1. Configuración de los experimentos

El número de muestras obtenidas para la paprika es de 15 y para la alcachofa 14, esto se debe a que el nitrógeno es volátil y difícil de medir, por tanto, la frecuencia con que se toma las muestras debe ser en espacios de tiempo donde se pueda notar la variabilidad de sus valores, por lo que nos encontramos ante un problema de pocas muestras. Tal como se ha descrito en la Subsección 2.2.3, existen varios procedimientos para lidiar con este tipo de problemas. En este caso, la propuesta del estimador es un modelo simple cuyos pesos se entrenan utilizando la Ecuación 4.7.

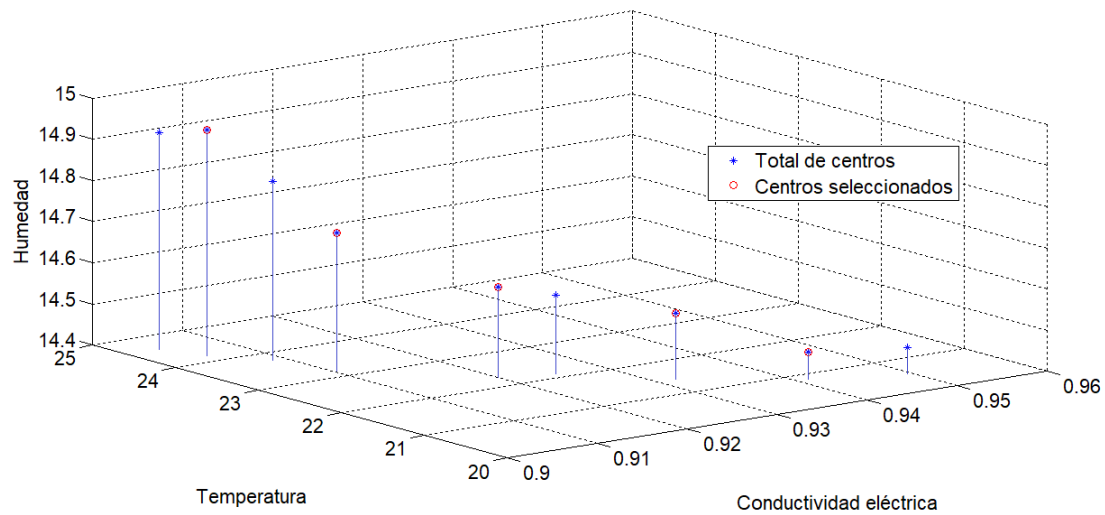
Antes de entrenar los pesos, las bases de datos se particionan de modo que el 70 % es para entrenar y el 30 % para probar el modelo entrenado. Con los datos de entrenamiento, primero se eligen los centroides de las funciones base utilizando la Ecuación 4.9. El valor de R se fija a la unidad luego de notar que en el diagrama de dispersión existe una tendencia a agruparse. Este comportamiento se muestra en la Figura 5.1a para el caso de la paprika y la Figura 5.1b para el caso de la alcachofa. Luego de elegir los centroides, se procede a calcular el valor del parámetro de dispersión de las funciones base de acuerdo con lo que indica la Ecuación 4.10, donde el valor de r se explora utilizando el procedimiento de LOOCV, en un intervalo que va desde 1.25 hasta 1000 espaciado en pasos logarítmicos.

5.2. Experimentos

Se ha desarrollado experimentos con diferentes números de variables de entrada. El primero cuando la variable de entrada es la Conductividad Eléctrica (CE), segundo cuando las variables de entrada son Temperatura (T) y CE, la tercera cuando las instancias son la CE y Humedad (H), finalmente, se utilizan las tres variables medidas de CE, T y H.



(a)

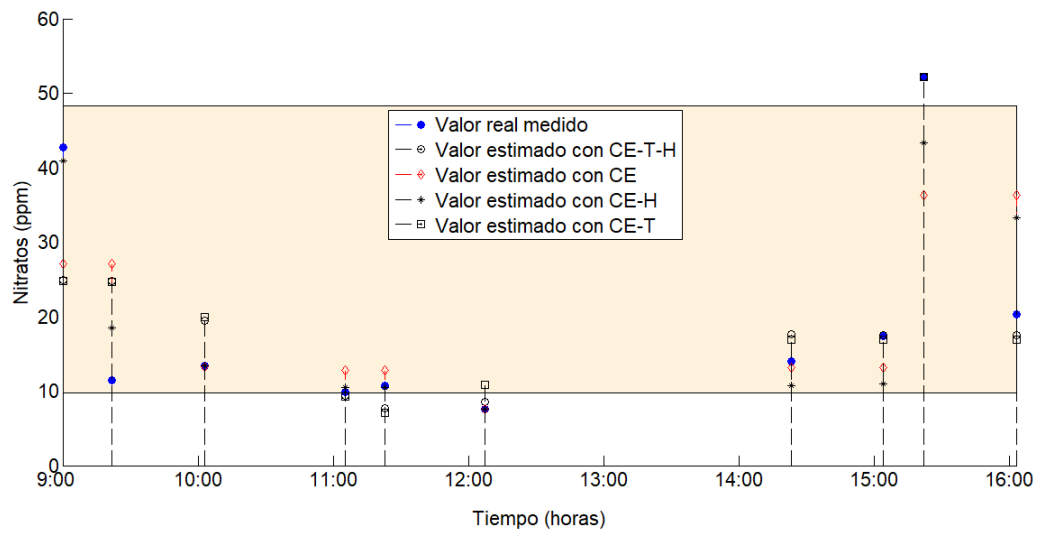


(b)

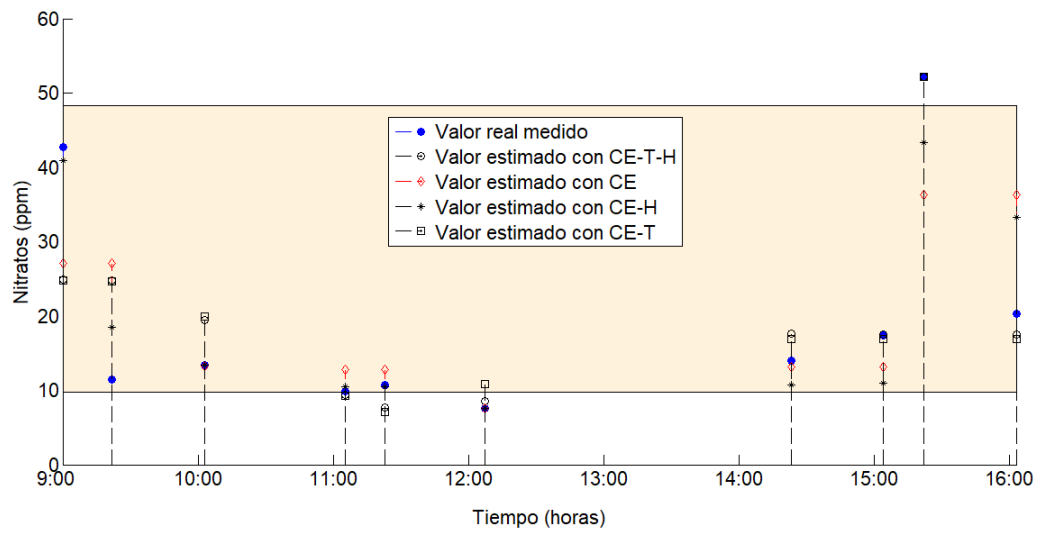
Figura 5.1: Diagrama de dispersión de los centros seleccionados para el modelo estimador de: (a) paprika y (b) alcachofa.

El diagrama de los experimentos para el entrenamiento del modelo máquina con las diferentes variables sin normalizar y normalizados se observan en la Figura 5.2 para la paprika, donde el área sombreada indica los rangos de los niveles de nitrógeno permitidos para el cultivo, según las recomendaciones dadas en la Subsubsección 2.1.2.2.

Para el cultivo de alcachofa el diagrama de los experimentos de entrenamiento del modelo máquina con las diferentes variables sin normalizar y normalizados se obser-



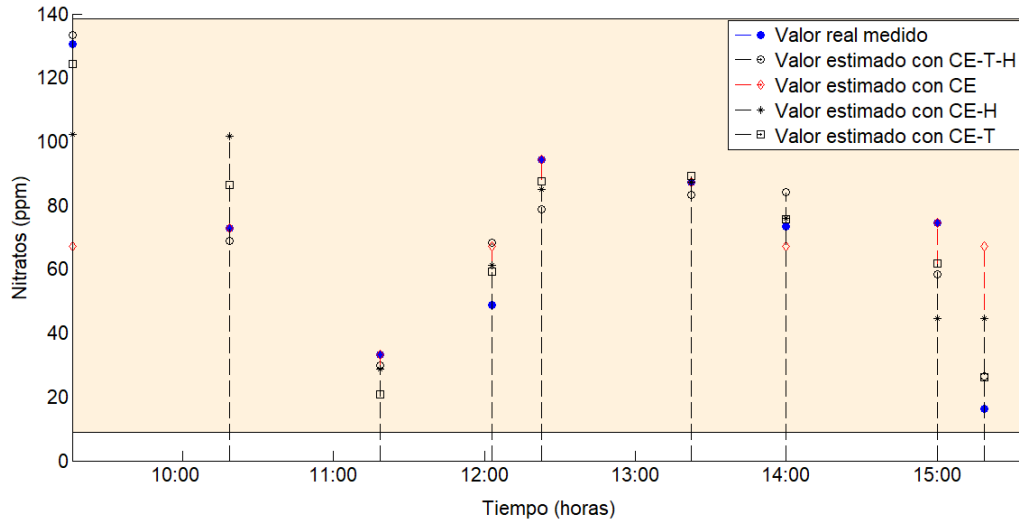
(a)



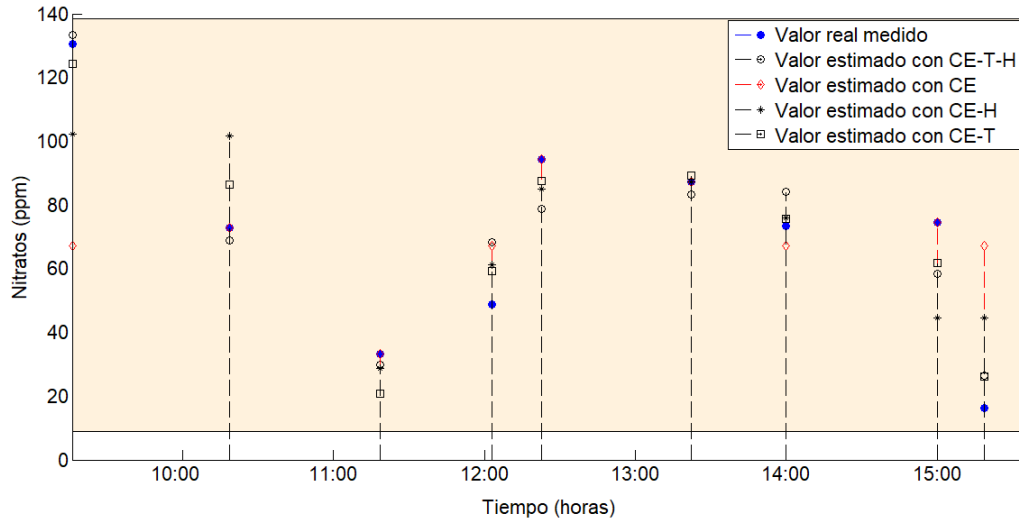
(b)

Figura 5.2: Valores del entrenamiento del modelo máquina para la paprika con diferentes variables de entrada. **(a)**: con las variables de entrada sin normalizar y **(b)**: con las variables de entrada normalizadas.

van en la Figura 5.3.



(a)



(b)

Figura 5.3: Valores del entrenamiento del modelo máquina para la alcachofa con diferentes variables de entrada. **(a):** con las variables de entrada sin normalizar y **(b):** con las variables de entrada normalizadas.

5.3. Resultados

Al realizar los experimentos con los datos de entrenamiento para el modelo máquina, se obtuvo los promedios del error absoluto y desviación estándar que se detallan en la Tabla 5.1.

	$ N - \hat{N} $			
	Sin normalizar			
	CE	CE-T	CE-H	CE-T-H
Paprika	$7,36 \pm 1,87 \times 10^{-15}$	$4,81 \pm 9,36 \times 10^{-16}$	$4,14 \pm 9,36 \times 10^{-16}$	$4,80 \pm 9,36 \times 10^{-16}$
Alcachofa	$15,41 \pm 3,74 \times 10^{-15}$	$8,49 \pm 3,74 \times 10^{-15}$	$16,02 \pm 3,74 \times 10^{-15}$	$9,54 \pm 1,87 \times 10^{-15}$
	Normalizado			
	CE	CE-T	CE-H	CE-T-H
Paprika	$7,36 \pm 1,87 \times 10^{-15}$	$4,81 \pm 9,36 \times 10^{-16}$	$4,14 \pm 9,36 \times 10^{-16}$	$4,80 \pm 9,36 \times 10^{-16}$
Alcachofa	$15,41 \pm 3,74 \times 10^{-15}$	$8,49 \pm 3,74 \times 10^{-15}$	$16,02 \pm 3,74 \times 10^{-15}$	$9,54 \pm 1,87 \times 10^{-15}$

Tabla 5.1: Promedio del valor absoluto del error y desviación estándar del estimador en el entrenamiento.

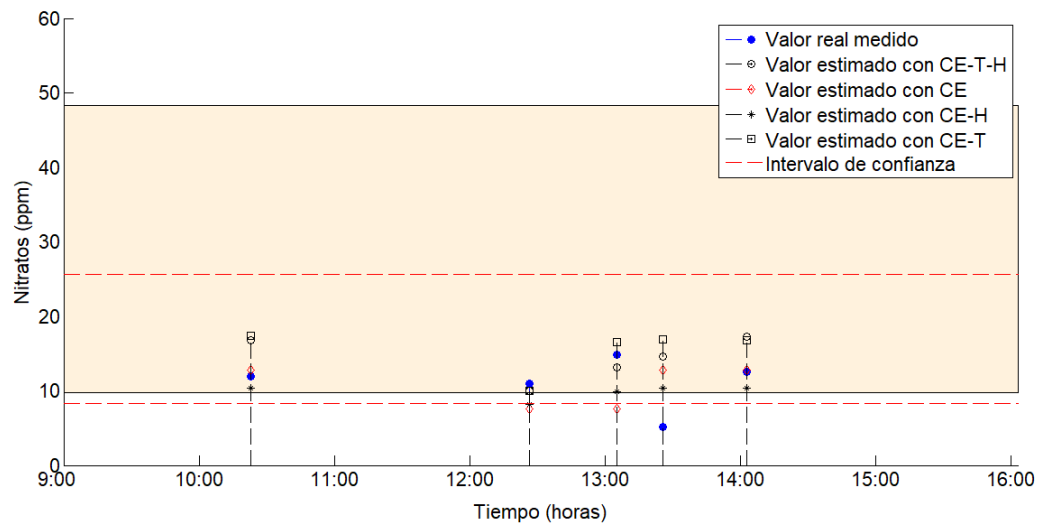
Para determinar el funcionamiento del modelo máquina, se hicieron pruebas de estimación del comportamiento de los niveles de nitrógeno con nuevos datos, no vistos en el entrenamiento. Los resultados del estimador con diferentes variables de entrada (CE-T-H) sin normalizar y normalizados para el cultivo de paprika se muestran en la Figura 5.4, donde las líneas entrecortadas horizontales indican los límites del intervalo de confianza al 99 %.

Los resultados del estimador para los datos de prueba con diferentes variables de entrada (CE-T-H) sin normalizar y normalizados para el cultivo de la alcachofa se muestran en la Figura 5.5, donde las líneas entrecortadas horizontales también indican los límites del intervalo de confianza al 99 %.

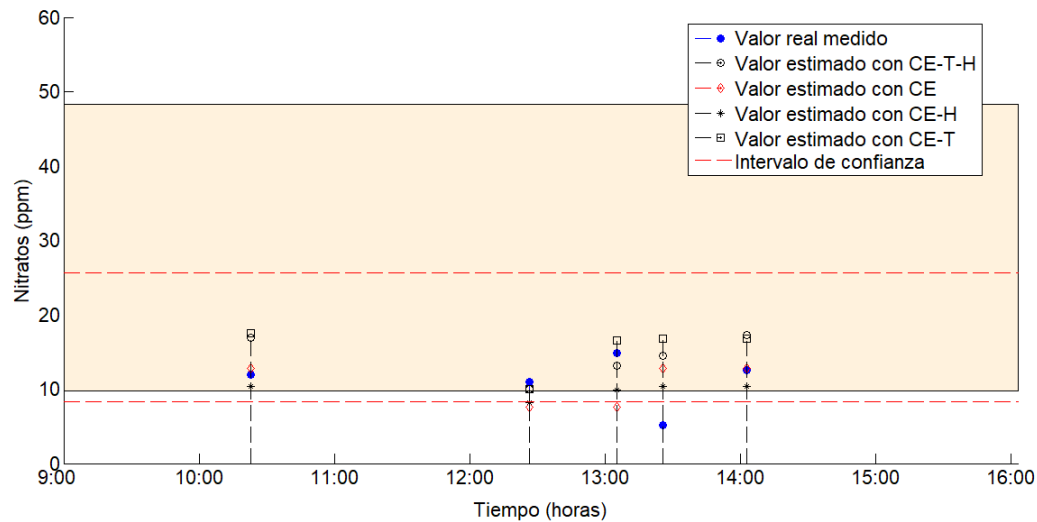
El cálculo del promedio del valor absoluto del error, para los resultados del estimador con los datos de prueba se muestran en la Tabla 5.2

5.4. Discusión

En la Tabla 5.2 se representa los resultados del estimador con los datos de prueba para una variable (CE), dos variables (CE-T, CE-H) y tres variables (CE-T-H) de entrada. En cada caso, se muestran los valores promedios luego de calcular el valor absoluto de la diferencia entre el valor real y el valor estimado luego de hacer 10 repeticiones. Todos los valores se han calculado bajo dos situaciones: cuando los datos están sin normalizar y cuando se normalizan a media cero y varianza unidad. A la vista de los resultados, tal como se esperaba, no son sensibles a la normalización. Sin embargo, es importante notar que el valor promedio del valor absoluto del error no indica cuál de los estimadores tiene mejores prestaciones, dado que no corresponden a



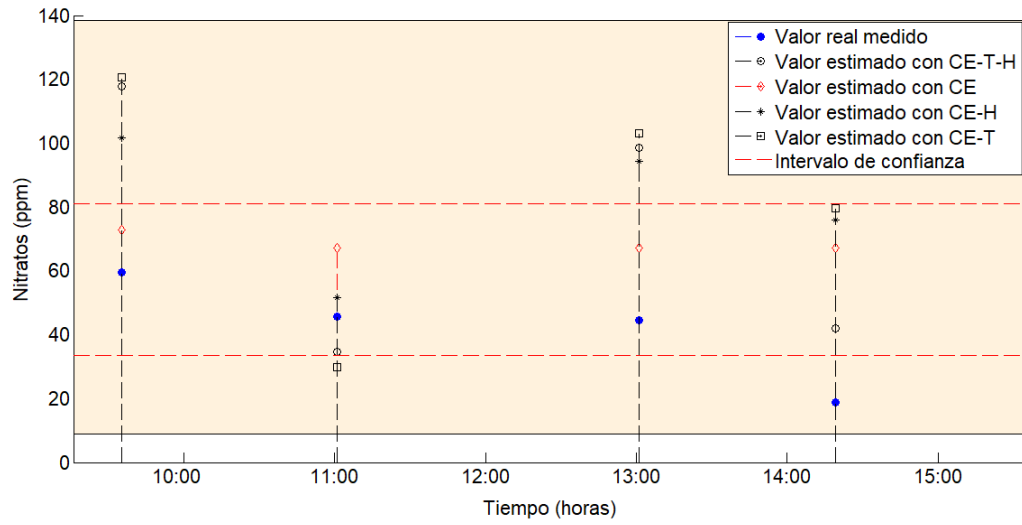
(a)



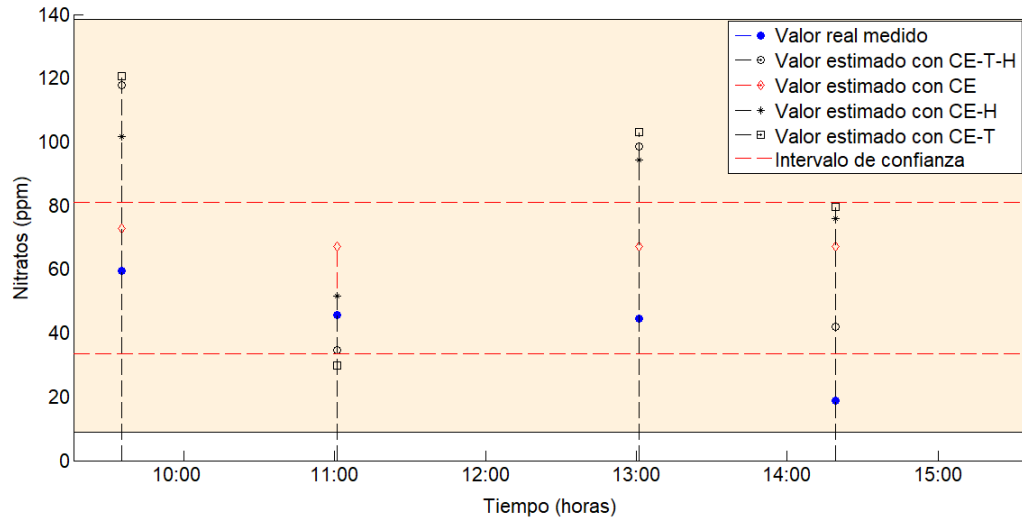
(b)

Figura 5.4: Resultados del estimador del nivel de nitrógeno con los datos de prueba para diferentes parámetros de entrada para el cultivo de paprika. **(a):** con parámetros sin normalizar y **(b):** con parámetros normalizados.

estimaciones puntuales, sino que nuestro objetivo es estimar dentro de un intervalo de confianza. En la Figura 5.4 y la Figura 5.5 se muestran los resultados de las estimaciones de los valores de nitrógeno y los rangos permitidos de cantidad de nitrógeno en suelo en ppm. También se indica los intervalos de confianza al 99 %. En el caso de la paprika teniendo en cuenta el intervalo de confianza, podemos decir que



(a)



(b)

Figura 5.5: Resultados del estimador del nivel de nitrógeno con los datos de prueba para diferentes parámetros de entrada para el cultivo de alcachofa. (a): con parámetros sin normalizar y (b): con parámetros normalizados.

el estimador entrenado sólo con la variable CE entrega resultados donde la mayoría de los puntos (niveles de nitrógeno) están dentro del intervalo de confianza, cuando se consideran dos variables de entrada, en el caso de CE-H la estimación mejora pero tiende a que los puntos estén fuera del intervalo, en el caso de CE-T todos los puntos son estimados dentro del intervalo, cuando se utilizan las tres variables

	$ N - \hat{N} $			
	Sin normalizar			
	CE	CE-T	CE-H	CE-T-H
Paprika	$3,91 \pm 1,87 \times 10^{-15}$	$4,33 \pm 9,36 \times 10^{-16}$	$3,40 \pm 9,36 \times 10^{-16}$	$4,32 \pm 9,36 \times 10^{-16}$
Alcachofa	$38,31 \pm 3,74 \times 10^{-15}$	$35,29 \pm 3,74 \times 10^{-15}$	$40,83 \pm 1,87 \times 10^{-15}$	$35,34 \pm 1,87 \times 10^{-15}$
	Normalizado			
	CE	CE-T	CE-H	CE-T-H
Paprika	$3,91 \pm 1,87 \times 10^{-15}$	$4,33 \pm 9,36 \times 10^{-16}$	$3,40 \pm 9,36 \times 10^{-16}$	$4,32 \pm 9,36 \times 10^{-16}$
Alcachofa	$38,31 \pm 3,74 \times 10^{-15}$	$35,29 \pm 3,74 \times 10^{-15}$	$40,83 \pm 1,87 \times 10^{-15}$	$35,34 \pm 1,87 \times 10^{-15}$

Tabla 5.2: Promedio del valor absoluto del error y desviación estándar de los niveles estimados para los datos de prueba.

CE-T-H todos los puntos estimados también se encuentran dentro del intervalo de confianza. Para el caso de la alcachofa, cuando el estimador es entrenado con una sola variable CE todos los puntos son estimados dentro del intervalo de confianza, cuando se entrena con las variables CE-H algunos puntos quedan fuera del intervalo, sin embargo, cuando se entrena con las dos variables CE-T o con las tres variables CE-T-H los puntos estimados tienden a estar dentro del intervalo de confianza.

6 Conclusiones y trabajos futuros

6.1. Conclusiones

- El nitrógeno es uno de los macronutrientes esenciales a la hora de conocer el estado de la planta y el suelo agrícola. Conocer si este está disponible para que sea aprovechado por el cultivo es importante dado que este se puede percolar si no se tiene cuidado a la hora de otorgar riego al campo. Por ello se ha desarrollado un estimador máquina que permite conocer los niveles del nitrógeno en el suelo agrícola de forma automática a partir de información de las condiciones ambientales del suelo, es decir, la temperatura, conductividad eléctrica y humedad. Estas variables en la mayoría de casos se miden en tiempo real. Las muestras utilizadas para el diseño del estimador corresponden a intervalos de tiempo de media hora porque las variaciones de los macro y micronutrientes no presentan variaciones en intervalos cortos de tiempo. En un terreno arenoso la variación de los niveles de nitrógeno tiene un comportamiento no uniforme ya que su textura presenta mayor porosidad lo que origina una mayor percolación y movimiento de los nutrientes, y en un terreno franco arenoso se presentan variaciones suaves por tener una estructura más equilibrada y de mayor enganche.
- Los resultados obtenidos del estimador muestran que la variable de entrada más importante para el entrenamiento del modelo máquina es la conductividad eléctrica, ya que es una variable que está íntimamente ligada con los minerales del suelo por lo que tiene un alto impacto con el comportamiento de los niveles del nitrógeno. Cuando se considera dos variables para el entrenamiento los mejores resultados son los de CE-T ya que los niveles estimados tienden a estar dentro del intervalo de confianza, sin embargo, el estimador entrenado con CE-H tiende a que los niveles estimados de nitrógeno salgan del intervalo de confianza ya que la humedad no aporta a la mejora del estimador. Considerar las tres variables de entrada CE-T-H en el entrenamiento también tiende a que los niveles de nitrógeno estimados estén dentro del intervalo de confianza en la mayoría de los puntos.
- Se ha diseñado un estimador no lineal con funciones base gaussianas porque el comportamiento del nitrógeno es no lineal. Dado que se estaba frente a un problema de pocas muestras, el diseño no contempla estimaciones puntuales, si no, intervalos de confianza. A la vista de los resultados, el estimador diseñado otorga resultados razonables dentro de un intervalo de confianza del 99 %

cuando se tiene como parámetros de entrada a las variables CE-T y CE-T-H para la paprika y CE para la alcachofa.

- Con los resultados obtenidos en las pruebas del modelo máquina podemos decir que el estimador cubre las expectativas de un estimador eficiente cuando al entrenar el modelo con una sola variable se considera CE, con dos variables se considera CE-T y con tres variables se considera CE-T-H, ya que los resultados de los niveles de nitrógeno estimados están en su mayoría dentro del intervalo de confianza dados por el rango de los niveles permitidos para el cultivo.

6.2. Publicaciones generadas

Es importante dar a conocer que este trabajo de tesis aparece disponible y publicada por la Conferencia ANDESCON, 2016 IEEE, [Sánchez Mora et al., 2016] con el título “A nonlinear model to estimate nitrogen level in agricultural soil using Gaussian kernels”.

6.3. Trabajos futuros

De los experimentos desarrollados se ha podido notar que el entrenamiento del modelo podría mejorar con la obtención de un mayor número de muestras para entrenamiento y para pruebas.

Los métodos kernel presentan buen comportamiento en series de tiempo con variaciones locales abruptas por ello un trabajo futuro puede enfocarse en probar la selección de centroides y dispersiones para mejorar las estimaciones.

Reconocimientos

Gracias

A mi querido padre Alberto Sánchez, por darme la oportunidad de crecer y poder desarrollarme profesionalmente. Gracias por todo tu esfuerzo y por dar lo mejor de ti.

Por ser tu mayor anhelo y orgullo que me convirtiera en una ingeniera como tú lo fuiste, es que te dedico este trabajo de tesis.

Gracias

A mi querida madre Ruth Mora, por enseñarme a perseverar y a no rendirme ante los problemas que se presentaban a lo largo de la carrera. Gracias por confiar en mí y estar a mi lado para apoyarme.

Gracias

A mi asesor de tesis: Dr Efraín Mayhua López por guiarme y compartir sus conocimientos y experiencias durante mi estudio de tesis. Gracias por ser más que un asesor, por ser un amigo quien me ayudó a no rendirme dándome palabras de ánimo y fortaleza. Gracias por brindarme cada minuto de su tiempo, los que en realidad se convirtieron en largas horas de trabajo.

Bibliografía

- [Cavigelli et al., 2004] Cavigelli, M. A., Mosier, A. R., Syers, J. K., and Freney, J. R. (2004). *Agriculture and the Nitrogen Cycle*.
- [Decagon, 2016] Decagon (2016). *5TE Water Content, EC and Temperature Sensor*. Decagon Devices, Inc.
- [Duda et al., 1973] Duda, R. O., Hart, P. E., and Stork, D. G. (1973). *Pattern Classification -Second edition*.
- [Gaikwad and Galande, 2015] Gaikwad, S. and Galande, S. (2015). Measurement of npk, temperature, moisture, humidity using wsn. *Journal of Engineering Research and Applications*.
- [Germán, 2013] Germán, V. (2013). Etapas de la toma de muestra para análisis de suelo. Technical report, Bolsa de Comercio de Rosario.
- [Hassan, 2011] Hassan, S. A. (2011). Effects of excess nitrogen on growth, flowering and fruit set of bell pepper. *IDEALS*.
- [Jeo, 1997] Jeo, T. (1997). Environment/ecology. <http://www.madsci.org/posts/archives/1997-03/857507931.En.r.html>.
- [Jiang et al., 2008] Jiang, Y., Li, M., and Zhou, Z. H. (2008). Mining extremely small data sets with application to software reuse. *Software: Practice and Experience*.
- [Johnson et al., 2005] Johnson, C., Albrecht, G., Ketterings, Q., Beckman, J., and Stockin, K. (2005). Nitrogen basics -the nitrogen cycle. Technical report, Cornell University Cooperative Extension.
- [Kissel and Sonon, 2008] Kissel, D. E. and Sonon, L. (2008). Soil test handbook for georgia. Technical report, College of Agricultural & Environmental Sciences - University of Georgia.
- [Labconco, 1998] Labconco, C. (1998). A guide to kjeldahl nitrogen determination methods and apparatus. *Labconco Corporation: Houston, TX, USA*.
- [Malghani et al., 2010] Malghani, A. L., Malik, A. U., Sattar, A., Hussain, F., Abbas, G., and Hussain, J. (2010). Response of growth and yield of wheat to npk fertilizer. *Sci. Int.(Lahore)*, 24(2):185–189.
- [Mayhua-López, 2012] Mayhua-López, E. (2012). Real adaboost with gate controlled fusion. *IEEE*.

- [Perdomo and Barbazán, 2011] Perdomo, C. and Barbazán, M. (2011). Nitrógeno. Technical report, Facultad de Agronomía Universidad de la Republica, Montevideo.
- [Rogers and Girolami, 2012] Rogers, S. and Girolami, M. (2012). *A First Course in Machine Learning*. CRC Press.
- [Shaw, 2014] Shaw, R. (2014). In-situ monitoring of soil nitrate in real time using ion-selective electrodes.
- [Shaw, 2015a] Shaw, R. (2015a). *Developing in-situ and real-time methods of soil nitrogen determination*. PhD thesis, Bangor University.
- [Shaw, 2015b] Shaw, R. (2015b). Developing in-situ and real-time methods of soil nitrogen determination. *Agriculture & Horticulture Development Board*.
- [Shaw et al., 2013] Shaw, R., Prysor, A., Miller, A., and Jones, D. (2013). Assessing the potential for ion selective electrodes and dual wavelength uv spectroscopy as a rapid on-farm measurement of soil nitrate concentration. *Agriculture*.
- [Shaw et al., 2014] Shaw, R., Williams, A., and Jones, D. (2014). Assessing soil nitrogen availability using microdialysis-derived diffusive flux measurements. *Soil Science Society of America Journal. Digital Library*.
- [Sánchez Mora et al., 2016] Sánchez Mora, K., Mayhua López, E., and Zuñiga Gutierrez, M. (2016). A nonlinear model to estimate nitrogen level in agricultural soil using gaussian kernels. *ANDESCON, IEEE*.
- [Thompson et al., 2002] Thompson, M., Owen, L., Wilkinson, K., Wood, R., and Damant, A. (2002). A comparison of the kjeldahl and dumas methods for the determination of protein in foods, using data from a proficiency testing scheme. *Analyst*, 127(12):1666–1668.
- [Young-Sup and Sung-Yang, 1996] Young-Sup, H. and Sung-Yang, B. (1996). An efficient method to construct a radial basis function neural network classifier and its application to unconstrained handwritten digit recognition. In *Pattern Recognition*.

Nomenclatura

H_2O	Agua
K_2SO_4	Sulfato de potasio
AA	Aminoácidos
APC-III	Adaptive Patern Classifier III
CE	Conductividad eléctrica
DSM	Direct Soil Measurement
H	Humedad
ISEs	Ion-selective electrodes
K	Potasio
KCl	cloruro de potasio
LOOCV	Leave-one-out cross validation
ML	Machine Learning
N	Nitrógeno
NPK	Nitrógeno - Potasio - Fósforo
P	Fósforo
PA	Precision agriculture
PPy	Pre-synthesized polypyrrole
PVC	Policloruro de vinilo
T	Temperatura